

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:41:13 ; Search time 22 Seconds
(without alignments)
2119.327 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 485

Sequence: 1 MITRETLKSLPANVQAPPYD.....KVIIVLRKIIREKGVKAAIPA 485

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	2.7	494	2 AG3103	mannitol 2-dehydro
2	13	2.7	525	2 C98183	mannitol 2-dehydro
3	10	2.1	486	2 S56548	fructuronate reduc
4	10	2.1	486	2 B91289	D-mannosate oxidor
5	10	2.1	486	2 E86130	mannonate oxidore
6	10	2.1	491	2 D83353	mannitol dehydroge
7	9	1.9	486	2 A64909	probable fructuron
8	9	1.9	486	2 G90897	probable oxidoredu
9	9	1.9	486	2 H85719	probable oxidoredu
10	9	1.9	488	2 AD0679	fructuronate reduc
11	8	1.6	244	2 B97627	hypothetical prote
12	8	1.6	244	2 AD2850	hypothetical prote
13	8	1.6	253	2 S76761	hypothetical prote
14	8	1.6	260	2 I40886	glycine hydroxymet
15	8	1.6	384	2 S73429	glycerol-3-phospat
16	8	1.6	409	2 T36940	probable oxidoredu
17	8	1.6	463	2 AF0283	probable mannitol
18	8	1.6	580	2 A86834	1-deoxyxylulose-5-
19	7	1.4	42	2 T13269	hypothetical prote
20	7	1.4	72	2 H81016	hypothetical prote
21	7	1.4	77	2 A47070	ferrochelatase [EC
22	7	1.4	130	2 A18175	hypothetical prote
23	7	1.4	142	2 S77423	hypothetical prote
24	7	1.4	142	2 C87255	transcriptional regu
25	7	1.4	149	2 A90524	hypothetical prote
26	7	1.4	154	2 B69840	conserved hypothet
27	7	1.4	166	2 D83953	hypothetical prote
28	7	1.4	171	2 C64247	conserved hypothet
29	7	1.4	172	2 T45196	probable transcrip

ALIGNMENTS

RESULT 1

AG3103

mannitol 2-dehydrogenase mtlK [imported] - Agrobacterium tumefaciens (strain C58, Dup
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AG3103

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193

A:Accession: AG3103

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-494 <KUR>

A:Cross-references: GB:AE008689; PIDN:AA45245.1; PID:g17742929; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: mtlK

A:Map position: linear chromosome

C:Superfamily: conserved hypothetical protein YEL070w

Query Match 2.7%; Score 13; DB 2; Length 494;

Best Local Similarity 100.0%; Pred.No. 8.3e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 IRIVSWTITEGGY 135

Db 124 IRIVSWTITEGGY 136

RESULT 2

C98183

mannitol 2-dehydrogenase (mdh) [imported] - Agrobacterium tumefaciens (strain C58, Ce
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002

C:Accession: C98183

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: C98183

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-525 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK88989.1; PID:g15158773; GSPDB:GN00170

C:Genetics:

A:Gene: AGR_L_830

A;Residues: 1-486 <HAY>
A;Cross-references: GB:BA000007; PIDN:BA38705.1; PID:gl3364760; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs5282
C;Superfamily: conserved hypothetical protein YEL070w

Query Match 2.1%; Score 10; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 KAFTVMSCDN 190
|||||
DB 178 KAFTVMSCDN 187
|||||

RESULT 5
E86130
mannanase oxidoreductase [imported] - Escherichia coli (strain O157:H7, substrain EDL
C;Species: Escherichia coli
C;Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C;Accession: E86130
R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Soderbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda
Nature 409, 529-533, 2001
A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A;Reference number: A85480; MUID:21074935; PMID:11206551
A;Accession: E86130
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-486 <STO>
A;Cross-references: GB:AE005174; NID:gi2519333; PIDN:AAG59505.1; GSPDB:GN00145; UMGp
A;Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: uuxB
C;Superfamily: conserved hypothetical protein YEL070w

Query Match 2.1%; Score 10; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 KAFTVMSCDN 190
|||||
DB 178 KAFTVMSCDN 187
|||||

RESULT 6
D83353
mannitol dehydrogenase PA2342 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: D83353
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; L
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: D83353
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-491 <STO>
A;Cross-references: GB:AE004660; GB:AE004091; NID:g9948372; PIDN:AAG05730.1; GSPDB:GN
A;Experimental source: strain PA01
C:Genetics:
A:Gene: mtID; PA2342
C;Superfamily: conserved hypothetical protein YEL070w

Query Match 2.1%; Score 10; DB 2; Length 491;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AFTVMSCDN 191

Db 183 AFTVMSCDNL 192
 |||||
 RESULT 7
 A64909
 probable fructuronate reductase (EC 1.1.1.57) ydfI - Escherichia coli (strain K-12)
 N:Alternate names: D-mannonate oxidoreductase
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: A64909
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC
 A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: A64909
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-486 <BLAT>
 A:Cross-references: GB:AE000251; GB:U00096; NID:g1787814; PIDN:AAC74615.1; PID:g1787823;
 A:Experimental source: strain K-12, substrain MGI655
 C:Genetics:
 A:Gene: ydfI
 C:Function:
 A:Description: reversible reaction of D-mannonate and NAD(+) to D-fructuronate and NADH
 A:Pathway: D-glucuronate catabolism
 A:Note: part of the interconversion of D-fructuronate to D-2-keto-3-deoxy-D-gluconate
 A:Note: reduces also D-tagaturonate
 C:Superfamily: conserved hypothetical protein YEL070w
 C:Keywords: carbon-oxygen lyase; hydro-lyase; NAD; oxidoreductase
 Query Match 1.9%; Score 9; DB 2; Length 486;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 182 AFTVMSCDN 190
 |||||
 Db 176 AFTVMSCDN 184
 |||||
 RESULT 8
 G90897
 probable oxidoreductase ECs2151 [imported] - Escherichia coli (strain O157:H7, substrain
 C:Species: Escherichia coli
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: G90897
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: G90897
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-486 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA035574.1; PID:g13361617; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: ECs2151
 C:Superfamily: conserved hypothetical protein YEL070w
 Query Match 1.9%; Score 9; DB 2; Length 486;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 182 AFTVMSCDN 190
 |||||
 Db 176 AFTVMSCDN 184
 |||||
 RESULT 9
 H85719
 probable oxidoreductase ydfI [imported] - Escherichia coli (strain O157:H7, substrain
 C:Species: Escherichia coli
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: H85719
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
 iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
 A:Reference number: H85480; MUID:21074935; PMID:11206551
 A:Accession: H85719
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-486 <STO>
 A:Cross-references: GB:AE005174; NID:g12515111; PIDN:AAG56220.1; GSPDB:GN00145; UWGP:
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: ydfI
 C:Superfamily: conserved hypothetical protein YEL070w
 Query Match 1.9%; Score 9; DB 2; Length 486;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 182 AFTVMSCDN 190
 |||||
 Db 176 AFTVMSCDN 184
 |||||
 RESULT 10
 AD0679
 fructuronate reductase (EC 1.1.1.57) - Salmonella enterica subsp. enterica serovar Ty
 C:Species: Salmonella enterica subsp. enterica serovar Typhi
 A:Note: this species has also been called Salmonella typhi
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AD0679
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
 S.; Moule, S.; O'Gaora, P.
 Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
 A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se
 A:Reference number: AB0502; PMID:11677608
 A:Accession: AD0679
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-488 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD01805.1; PID:g16502650; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY1553
 C:Superfamily: conserved hypothetical protein YEL070w
 C:Keywords: oxidoreductase
 Query Match 1.9%; Score 9; DB 2; Length 488;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 182 AFTVMSCDN 190
 |||||
 Db 176 AFTVMSCDN 184
 |||||
 RESULT 11
 B97627
 hypothetical protein AGR_C_4053 [imported] - Agrobacterium tumefaciens (strain C58, C
 C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
 C:Accession: B97627
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldm
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
 Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
 A:Reference number: A97359; PMID:11743194
 A:Accession: B97627

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87971.1; PID:gl15157379; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4053
A:Map position: circular chromosome

Query Match 1.6%; Score 8; DB 2; Length 244;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 TLKSLPAN 13
|||||
Db 183 TLKSLPAN 190

RESULT 12
AD2850
hypothetical protein Atu2229 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AD2850
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S.; Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AD2850
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-244 <KUR>
A:Cross-references: GB:AE008688; PIDN:AA43218.1; PID:gl17740700; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu2229
A:Map position: circular chromosome

Query Match 1.6%; Score 8; DB 2; Length 244;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 TLKSLPAN 13
|||||
Db 183 TLKSLPAN 190

RESULT 13
S76761
hypothetical protein - Synecocystis sp. (strain PCC 6803)
C:Species: Synecocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, K.; DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S76761
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-253 <KAN>
A:Cross-references: EMBL:D90916; GB:AB001339; NID:gl653715; PIDN:BAAL18673.1; PID:gl65376
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Start codon: GTG

Query Match 1.6%; Score 8; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 437 LAKADDFE 444
|||||
Db 219 LAKADDFE 226

Search completed: May 8, 2003, 16:45:11
Job time : 34 secs

Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 VIPTLKAP 345
|||||
Db 2 VIPTLKAP 9

RESULT 14
I40886
glycine hydroxymethyltransferase (EC 2.1.1.2.1) - Corynebacterium sp. (fragment)
N:Alternate names: serine aldolase; serine hydroxymethylase; threonine aldolase
C:Species: Corynebacterium sp.
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
C:Accession: I40886
R:Chlumsky, L.J.; Zhang, L.; Jorns, M.S.
J. Biol. Chem. 270, 18252-18259, 1995
A:Title: Sequence analysis of sarcosine oxidase and nearby genes reveals homologies w
A:Reference number: A57385; MUID:95355441; PMID:7543100
A:Accession: I40886
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-260 <RES>
A:Cross-references: EMBL:U23955; NID:927587; PIDN:AA43458.1; PID:927588
C:Genetics:
A:Gene: glyA
C:Superfamily: glycine hydroxymethyltransferase
C:Keywords: phosphoprotein; pyridoxal phosphate; transferase
F:60/Binding site: Pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 260;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 AEIAKKLN 244
|||||
Db 75 AEIAKKLN 82

RESULT 15
S73429
glycerol-3-phosphate dehydrogenase glpD, aerob induced - Mycoplasma pneumoniae (strain
N:Alternate names: hypothetical protein D09_orf384
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 19-May-2000
C:Accession: S73429
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumonia
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73429
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-384 <HIM>
A:Cross-references: EMBL:AE000012; GB:U00089; NID:gl673755; PIDN:AA895751.1; PID:gl67
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: glpD
A:Genetic code: SCS3
C:Superfamily: Escherichia coli hypothetical protein ygaF

Query Match 1.6%; Score 8; DB 2; Length 384;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 437 LAKADDFE 444
|||||
Db 219 LAKADDFE 226

Search completed: May 8, 2003, 16:45:11
Job time : 34 secs

THIS PAGE BLANK (USPTO)


```
RP SEQUENCE FROM N.A.
RC STRAIN-SI-4;
RA Schneider K.-H., Giffhorn F.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-fructose + NADH.
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AF018073; AAC45771.1; -
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLHDHGRNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT INIT-MET 0
FT NP_BIND 18 29 NAD (BY SIMILARITY).
SQ SEQUENCE 476 AA; 51406 MW; A58247D6C7CCAF29 CRC64;
Query Match 2.3%; Score 11; DB 1; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.0064;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 223 FPNQVMDRITP 233
DB 211 FPNQVMDRITP 221
RESULT 3
UXUB_ECOLI STANDARD; PRT; 486 AA.
AC P39160;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE D-mannate oxidoreductase (EC 1.1.1.57) (Fructuronate reductase).
GN UXUB OR B4323.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes."
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA Mizobuchi K.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: D-mannate + NAD(+) = D-fructuronate + NADH.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
```

```
DR EMBL; U14003; AAA97219.1; -
DR EMBL; AE000503; AAC77279.1; -
DR EMBL; D13329; BAA02591.1; -
DR EcoGene; EG20248; uxuB.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLHDHGRNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD. Complete proteome.
FT NP_BIND 25 36 NAD (BY SIMILARITY).
SQ SEQUENCE 486 AA; 53580 MW; 7835638682AFB87 CRC64;
Query Match 2.1%; Score 10; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 KAFVMSCDN 190
DB 178 KAFVMSCDN 187
RESULT 4
YDFI_ECOLI STANDARD; PRT; 486 AA.
ID YDFI_ECOLI
AC P77260;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical oxidoreductase ydfi (EC 1.-.-.-).
GN YDFI OR B1542.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Alba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nashimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map."
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AE000251; AAC74615.1; -
DR EMBL; D90797; BAA15243.1; -
DR EMBL; D90798; BAA15247.1; -
DR EcoGene; EG13821; ydfi.
DR InterPro; IPR000669; Mannitol_dh.
```

```

DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Hypothetical protein; Oxidoreductase; NAD; Complete proteome.
FT NP_BIND 24 35 NAD (BY SIMILARITY).
SQ SEQUENCE 486 AA; 53685 MW; 96F3BF07AE114B70 CRC64;

Query Match 1.9%; Score 9; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 APTVMSCDN 190
DB 176 APTVMSCDN 184
|||||

* RESULT 5
GLYA_CORS1 STANDARD; PRT; 260 AA.
AC P50434;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase)
DE (SMT) (Fragment).
GN GLYA.
OS Corynebacterium sp. (strain P-1).
OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=69006;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=9535441; PubMed=7543100;
RA Chlumsky L.J., Zhang L., Jorns M.S.;
RT "Sequence analysis of sarcosine oxidase and nearby genes reveals
RT homologues with key enzymes of folate one-carbon metabolism.";
RL J. Biol. Chem. 270:18252-18259(1995).
CC -!- FUNCTION: INTERCONVERSION OF SERINE AND GLYCINE.
CC H(2)O = tetrahydrofolate + L-serine.
CC -!- Cofactor: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: KEY ENZYME IN THE BIOSYNTHESIS OF PURINES, LIPIDS,
CC HORMONES AND OTHER COMPONENTS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE SMT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U23955; AAC43458.1; -
DR HSP; P07511; 1C70.
DR InterPro; IPR001085; Gly_HyMettransf.
DR Pfam; PF00464; SMT; 1.
DR PROSITE; PS00096; SMT; 1.
KW Transferase; Pyridoxal phosphate; One-carbon metabolism.
FT NON_TER 1 1
FT BINDING 60 60 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 260 AA; 27578 MW; AD16613C7DAE467B CRC64;

Query Match 1.6%; Score 8; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 AEIAKKLN 244
DB 75 AEIAKKLN 82
|||||

RESULT 6
Y039_MYCPN STANDARD; PRT; 384 AA.
AC P75063;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG039 homolog (D09_orf384).
GN MPN051 OR MP103.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE000012; AAB95751.1; -
DR InterPro; IPR000927; D_aa_oxidase.
DR Pfam; PF01266; DAO; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 384 AA; 42724 MW; DA3E128719EE1BCD CRC64;

Query Match 1.6%; Score 8; DB 1; Length 384;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 LAKADDFE 444
DB 219 LAKADDFE 226
|||||

RESULT 7
DALD_YERPE STANDARD; PRT; 463 AA.
AC P58709;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE D-arabinitol 4-dehydrogenase (EC 1.1.1.11).
GN DALD OR YPO2325.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).

```

CC -!- CATALYTIC ACTIVITY: D-arabinitol + NAD(+) = D-xylulose + NADH.
 CC -!- PATHWAY: D-arabinitol catabolism; first step.
 CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AJ414152; CAC91130.1; .
 CC InterPro: IPR000669; Mannitol_dh.
 CC Pfam: PF01232; Mannitol_dh_1.
 CC PROSITE: PS00974; MANNITOL_DHGENASE; FALSE_NEG.
 KW Oxidoreductase; NAD; Complete proteome.
 SQ SEQUENCE 463 AA; 51984 MW; D82E988685B5EDD CRC64;

 Query Match 1.6%; Score 8; DB 1; Length 463;
 Best Local Similarity 100.0%; Pred. No. 6.6; Mismatches 0; Indels 0; Gaps 0;
 Matches 8; Conservative 0;
 QY 188 CDNLRRHG 195
 DB 168 CDNLRRHG 175
 |||||

 RESULT 8
 HEMZ_YERPS
 ID HEMZ_YERPS STANDARD; PRT; 77 AA.
 AC Q05338;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Ferrohelatase (EC 4.99.1.1) (Protoheme ferro-lyase) (Heme synthetase)
 DE (Fragment).
 GN HEMH.
 OS Yersinia pseudotuberculosis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=633;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93186709; PubMed=8444803;
 RA Kessler A.C., Haase A., Reeves P.R.;
 RT "Molecular analysis of the 3,6-dideoxyhexose pathway genes of
 RT Yersinia pseudotuberculosis serogroup IIA.";
 RL J. Bacteriol. 175:1412-1422(1993).
 CC -!- FUNCTION: CATALYZES THE FERROUS INSERTION INTO PROTOPORPHYRIN IX.
 CC -!- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) = protoheme + 2 H(+).
 CC -!- PATHWAY: Protoheme biosynthesis; last step.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE FERROCHELATASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L01777; ABA9397.1; .
 CC PIR: A47070; A47070.
 CC InterPro: IPR001015; Ferrochelatase.
 CC Pfam: PF00762; Ferrochelatase; 1.
 CC ProDom: PD002792; Ferrochelatase; 1.
 CC PROSITE: PS00534; FERROCHELATASE; PARTIAL.
 KW Porphyrin biosynthesis; Heme biosynthesis; Lyase; Iron.
 FT NON_TER 1
 FT METAL 32 32 IRON (BY SIMILARITY).
 SQ SEQUENCE 77 AA; 8766 MW; 5E3D1BA7AFACDESE CRC64;

Query Match 1.4%; Score 7; DB 1; Length 77;
 Best Local Similarity 100.0%; Pred. No. 14; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;
 QY 5 ETLKSLP 11
 DB 6 ETLKSLP 12
 |||||

 RESULT 9
 PTP_ACIOJ
 ID PTP_ACIOJ STANDARD; PRT; 142 AA.
 AC O52787;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Low molecular weight protein-tyrosine-phosphatase ptp (EC 3.1.3.48).
 GN PTP.
 OS Acinetobacter johnsonii.
 OC Bacteria; Proteobacteria; gamma subdivision; Moraxellaceae;
 OC Acinetobacter.
 OX NCBI_TaxID=40214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98094281; PubMed=9434192;
 RA Grangeasse C., Doubilet P., Vaganay E., Vincent C., Deleage G.,
 RA Duclos B., Cozzone A.J.;
 RT "Characterization of a bacterial gene encoding an autophosphorylating
 RT protein tyrosine kinase.";
 RL Gene 204:259-265(1997).
 CC [2]
 CC CHARACTERIZATION, AND MUTAGENESIS OF CYS-10 AND ARG-16.
 CC MEDLINE=98239739; PubMed=9571056;
 CC Grangeasse C., Doubilet P., Vincent C., Vaganay E., Riberty M.,
 CC Duclos B., Cozzone A.J.;
 CC "Functional characterization of the low-molecular-mass
 CC phosphotyrosine-protein phosphatase of Acinetobacter johnsonii.";
 CC J. Mol. Biol. 278:339-347(1998).
 CC -!- FUNCTION: Dephosphorylates ptk. May be involved in the production
 CC and the transport of exopolysaccharides.
 CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2O) = protein
 CC tyrosine + phosphate.
 CC -!- ENZYME REGULATION: Inhibited by ammonium molybdate, sodium
 CC orthovanadate, N-ethylmaleimide and iodoacetic acid.
 CC -!- PATHWAY: Exopolysaccharide biosynthesis
 CC -!- SIMILARITY: BELONGS TO THE LOW MOLECULAR WEIGHT PHOSPHOTYROSINE
 CC PROTEIN PHOSPHATASES FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Y15162; CAA75430.1; .
 CC HSP; P24666; SPNT.
 CC InterPro: IPR000106; Low_mwt_PTPase.
 CC Pfam: PF01451; LMWPC; 1.
 CC SMART: SM00226; LMWPC; 1.
 CC Hydrolase; Exopolysaccharide synthesis.
 CC ACT_SITE 10 10 NUCLEOPHILE (BY SIMILARITY).
 CC ACT_SITE 15 15 BY SIMILARITY.
 CC ACT_SITE 115 115 PROTON DONOR (BY SIMILARITY).
 CC MUTAGEN 10 10 C->S: LOSS OF ACTIVITY.
 CC MUTAGEN 16 16 R->K: LOSS OF ACTIVITY.
 CC SEQUENCE 142 AA; 16215 MW; 62B53F3BDDBA5986 CRC64;

 Query Match 1.4%; Score 7; DB 1; Length 142;
 Best Local Similarity 100.0%; Pred. No. 23; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;

```

QY 239 IAKKLNA 245
DB 65 IAKKLNA 71

RESULT 10
Y428_MYCSE
ID Y428_MYCSE STANDARD; PRT: 171 AA.
AC P47667;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG428.
GN MG428.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Otterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL: U39725; AAC72449.1;
DR TIGR: MG428;
DR InterPro: IPR000792; HTH_LuxR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 171 AA; 20259 MW; 87FB346C334E5E72 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 171;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 ETAKKLN 244
DB 139 ETAKKLN 145

RESULT 11
Y428_MYCSE
ID Y428_MYCSE STANDARD; PRT: 172 AA.
AC P75169;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG428 homolog (Cl2_orf172).
GN MPN026 OR MP216.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma

```

```

pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

DR EMBL: AF000021; AAB95864.1;
DR InterPro: IPR000792; HTH_LuxR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 172 AA; 20473 MW; 93F6430F3BF8C4AB CRC64;

Query Match 1.4%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 238 ETAKKLN 244
DB 140 ETAKKLN 146

RESULT 12
YPIB_BACSU
ID YPIB_BACSU STANDARD; PRT: 179 AA.
AC P54390;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ypiB.
GN YPIB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / Marburg;
RX MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA Serror P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kdg loci cloned in a yeast artificial chromosome.";
RL Microbiology 142:2005-2016(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinol S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,

```

RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambui R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT *subtilis*.";
 RL Nature 390:249-256(1997).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L47709; AAC38433.1; -;
 DR EMBL; Z99115; CAB14174.1; -;
 DR Subtilist; BG11497; Yp1B.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 179 AA; 21364 MW; 46366E994DB709D9 CRC64;
 Query Match 1.4%; Score 7; DB 1; Length 179;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 46 EQILEHA 52
 DB 137 EQILEHA 143
 RESULT 13
 WCAF_ECOLI
 ID WCAF_ECOLI STANDARD; PRT; 182 AA.
 AC P71240; P76383;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative colanic acid biosynthesis acetyltransferase wcaf
 DE (EC 2.3.1.-).
 GN WCAF OR B2054.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=96326333; PubMed=8759852;
 RA Stevenson G., Andrianopoulos K., Hobbs M., Reeves P.R.;
 RT "Organization of the *Escherichia coli* K-12 gene cluster responsible
 RT for production of the extracellular polysaccharide colanic acid.";
 RL J. Bacteriol. 178:4885-4893(1996).
 RN [2]
 RN REVISIONS TO N-TERMINUS.
 RC STRAIN-K12;
 RA Reeves P.R.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97251358; PubMed=9097040;

RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
 RA Isino K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
 RA Saito N., Sempel G., Seki Y., Sivasundaram S., Tagami H.,
 RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
 RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome
 RT corresponding to the 40.1-50.0 min region on the linkage map.";
 RL DNA Res. 3:379-392(1996).
 CC -----
 CC -!- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF THE SLIME POLYSACCHARIDE
 CC COLANIC ACID.
 CC -----
 CC -!- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NODL FAMILY OF
 CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U38473; AAC77841.1; -;
 DR EMBL; AE000295; AAC75115.1; -;
 DR EMBL; D90843; BAA15910.1; -;
 DR EcoGene; EGI3574; wcaf.
 DR InterPro; IPR001451; Hexapep_transf.
 DR Pfam; PF00132; hexapep; 3.
 DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; 1.
 KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase; Repeat;
 KW Complete proteome.
 FT CONFLICT 50 F -> L (IN REF. 2).
 SQ SEQUENCE 182 AA; 19962 MW; 8AA5E672CA802442 CRC64;
 Query Match 1.4%; Score 7; DB 1; Length 182;
 Best Local Similarity 100.0%; Pred. No. 29;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 8 KSLPANV 14
 DB 159 KSLPANV 165
 RESULT 14
 DGC6_CHICK
 ID DGC6_CHICK STANDARD; PRT; 200 AA.
 AC O73770;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DGC6 protein (Digeorge syndrome critical region 6 homolog).
 GN DGC6.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN-Broiler; TISSUE=Heart;
 RA Hierck B.P., Poelmann R.E., Gittenberger-De Groot A.C.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO THE GONADAL FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF048985; AAC05142.1; -;

KW Coiled coil. 116 146 COILED COIL (POTENTIAL).
FT DOMAIN 176 198 COILED COIL (POTENTIAL).
SQ SEQUENCE 200 AA; 23242 MW; 7037BBF6F660E469 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 ELDQKVI 468
Db 146 ELDQKVI 152

RESULT 15

GTTL_DROER STANDARD; PRT; 200 AA.
AC P30104;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Glutathione S-transferase 1-1 (EC 2.5.1.18) (GST class-theta)
DE (Fragment).
GN GSTD1 OR GST1 OR GST.
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP SEQUENCE FROM N.A.
RA Hargis M.T., Cochrane B.J.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DDBJ databases.
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. THETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84576; -; NOT_ANNOTATED_CDS.
DR HSP; P30712; 1LJR.
DR FlyBase; FBgn0012266; Dere\Gstd1.
DR InterPro; IPR004046; GST_Cterm..
DR InterPro; IPR004045; GST_Nterm..
DR Pfam; PF00043; GST_C; 1..
DR Pfam; PF02798; GST_N; 1..
KW Transferase; Multigene family.
FT NON_TER 1
SQ SEQUENCE 200 AA; 22680 MW; 82052B2068FA4895 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 APADPEA 113
Db 113 APADPEA 119

Search completed: May 8, 2003, 16:42:56
Job time : 17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:40:54 ; Search time 89 Seconds
(without alignments)
1122.841 Million cell updates/sec

Title: US-09-926-163b-2
Perfect score: 485
Sequence: 1 MITRETLKSLPANVQAPPYD.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_ivirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	485	100.0	485	2 Q9KWR5	Q9kwr5 gluconobact
2	52	10.7	485	2 Q8VUU5	Q8vuu5 gluconobact
3	13	2.7	525	16 Q8U7J8	Q8u7j8 agrobacteri
4	11	2.3	492	16 Q98D06	Q98d06 rhizobium l
5	11	2.3	494	16 Q92N07	Q92n07 rhizobium m
6	10	2.1	486	16 Q8X516	Q8x516 escherichia
7	10	2.1	491	16 Q911D6	Q911d6 pseudomonas
8	10	2.1	493	2 Q88355	Q88355 pseudomonas
9	9	1.9	486	16 Q8X506	Q8x506 escherichia
10	9	1.9	488	16 Q8ZP18	Q8zpi18 salmonella
11	9	1.9	488	16 Q8Z626	Q8z626 salmonella
12	8	1.6	244	16 Q8UD98	Q8ud98 agrobacteri
13	8	1.6	253	16 P74566	P74566 synechocyst
14	8	1.6	261	5 Q9W439	Q9w439 drosophila
15	8	1.6	264	2 Q24792	Q24792 borrelia ga
16	8	1.6	301	16 Q8XUP2	Q8xup2 ralstonia s

17	1.6	374	16	Q9L1S8	Q9l1s8 streptomyce
18	1.6	409	16	Q9RJ07	Q9rj07 streptomyce
19	1.6	580	16	Q9CF08	Q9cf08 lactococcus
20	1.6	1006	10	Q9LT95	Q9lt95 arabidopsis
21	1.6	1030	3	Q9HFQ9	Q9hfg9 emericella
22	1.4	18	6	P79214	P79214 cryptotlagus
23	1.4	42	9	Q9AYV1	Q9ayv1 lactococcus
24	1.4	42	9	Q38332	Q38332 lactococcus
25	1.4	42	9	Q9B011	Q9b011 bacterioph
26	1.4	55	16	Q8XY21	Q8xy21 ralstonia s
27	1.4	72	16	Q9JR22	Q9jr22 neisseria m
28	1.4	106	10	Q94IU1	Q94iu1 fragaria ve
29	1.4	106	10	Q94IU3	Q94iu3 fragaria ve
30	1.4	110	17	Q9HKU9	Q9hku9 thermoplasm
31	1.4	128	10	Q93X92	Q93x92 lotus corni
32	1.4	130	16	Q8YZC5	Q8yzc5 anabaena sp
33	1.4	139	16	Q92RA2	Q92ra2 rhizobium m
34	1.4	142	16	P73486	P73486 synechocyst
35	1.4	142	16	Q9AC17	Q9ac17 caulobacter
36	1.4	147	17	Q976M0	Q976m0 sulfolobus
37	1.4	148	17	Q8U067	Q8u067 pyrococcus
38	1.4	149	9	O80313	O80313 bacterioph
39	1.4	149	16	Q98RB3	Q98rb3 mycoplasma
40	1.4	154	16	O06744	O06744 bacillus su
41	1.4	158	4	Q96D43	Q96d43 homo sapien
42	1.4	158	5	Q9NMK5	Q9nmk5 leishmania
43	1.4	164	16	Q98GU0	Q98gu0 rhizobium l
44	1.4	165	10	Q946W6	Q946w6 prunus pers
45	1.4	166	16	Q9KA62	Q9ka62 bacillus ha

ALIGNMENTS

RESULT 1

Q9KWR5 ID Q9KWR5 PRELIMINARY; PRT; 485 AA.
AC Q9KWR5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Sorbitol dehydrogenase.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G624;
RA Shibata T., Ichikawa C., Matsuura M., Takata Y., Noguchi Y., Saito Y.,
RA Yamashita M.;
RT "Cloning of a gene for D-Sorbitol dehydrogenase from Gluconobacter
oxydans G624 and expression of the gene in Pseudomonas putida
IFO3738.";
RT J. Biosci. Bioeng. 89:463-468(2000).
RL EMBL; AB028937; BAA99414.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLDHDRGNASE.
SQ SEQUENCE 485 AA; 53642 MW; A7C363D728EEA7DC CRC64;

Query Match 100.0%; Score 485; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MITRETLKSLPANVQAPPYDIDGIRPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
|||||

Db 1 MITRETLKSLPANVQAPPYDIDGIRPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
|||||

QY 61 GLUTGSDRSKKKAEFEKAQOCLYSLTETAPSGKSTVVMGALRDYLLAPADPAVLKHLVD 120
|||||

Db 61 GLUTGSDRSKKKAEFEKAQOCLYSLTETAPSGKSTVVMGALRDYLLAPADPAVLKHLVD 120
|||||

```
QY 121 PAIRIVSMITTEGGYNINETTGAFLDENAARKADLNKPEKSTVFYGVVEALRRRDAGG 180
Db 121 PAIRIVSMITTEGGYNINETTGAFLDENAARKADLNKPEKSTVFYGVVEALRRRDAGG 180
QY 181 KAFTVMSCDNLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWDRITPTVSAETA 240
Db 181 KAFTVMSCDNLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWDRITPTVSAETA 240
QY 241 KKLNAASGIDDDLPVLAEDPHQWLEDFQADGRPPLEKAGVOMGVDVTDWYVKIRMLNA 300
Db 241 KKLNAASGIDDDLPVLAEDPHQWLEDFQADGRPPLEKAGVOMGVDVTDWYVKIRMLNA 300
QY 301 GHVLMCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360
Db 301 GHVLMCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360
QY 361 FSNKMSDQTLRIASDCGSKVQVFTETVYRRATIEDKRDLSRIAFAFGIASYLEMLGRDEKG 420
Db 361 FSNKMSDQTLRIASDCGSKVQVFTETVYRRATIEDKRDLSRIAFAFGIASYLEMLGRDEKG 420
QY 421 GTYESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDQKVIIVLRKIIRKGVK 480
Db 421 GTYESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDQKVIIVLRKIIRKGVK 480
QY 481 AAIPA 485
Db 481 AAIPA 485

RESULT 2
Q8VUU5
ID Q8VUU5 PRELIMINARY; PRT: 485 AA.
AC Q8VUU5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE L-sorbose reductase.
GN SR.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFO3291;
RC MEDLINE=21650687; PubMed=11790761;
RA Shinjoh M., Tazoe M., Hoshino T.;
RT "NADPH-Dependent L-Sorbose Reductase Is Responsible for L-Sorbose
Assimilation in Gluconobacter suboxydans IFO 3291.";
RL J. Bacteriol. 184:861-863(2002).
DR EMBL; AB063188; BAB83933.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDDHGRNASE.
SQ SEQUENCE 485 AA; 53541 MW; 883EC4956B3C6CF0 CRC64;

Query Match 10.7%; Score 52; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.4e-44;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 MSCDNLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWDRITPTVSA 237
Db 186 MSCDNLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWDRITPTVSA 237

RESULT 3
Q8U7J8
ID Q8U7J8 PRELIMINARY; PRT: 525 AA.
AC Q8U7J8;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannitol 2-dehydrogenase.
```

```
GN MTLK OR ATU4451 OR AGR_L_830.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Stubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
Kutyavyn T., Levy R., Li M.-J., McClelland E., Palmieri A.,
Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RN [2]
RP Science 294:2317-2323(2001).
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
Houmiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58.";
RN [3]
RP Science 294:2323-2328(2001).
DR EMBL; AE009374; AAL45245.1; ALT_INIT.
DR EMBL; AE008240; AAK88989.1; -.
KW Complete proteome.
SQ SEQUENCE 525 AA; 57859 MW; F65432D35A4AFA6C CRC64;

Query Match 2.7%; Score 13; DB 16; Length 525;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 IRIVSMITTEGGY 135
Db 155 IRIVSMITTEGGY 167

RESULT 4
Q98D06
ID Q98D06 PRELIMINARY; PRT: 492 AA.
AC Q98D06;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannitol dehydrogenase.
GN MLL4920.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RC MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
Mesorhizobium loti.";
RN [2]
RP DNA Res. 7:331-338(2000).
DR EMBL; AP003005; BAB51465.1; -.
```

```

DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 492 AA; 53518 MW; 4D20E1136C3C1373 CRC64;

Query Match          2.3%; Score 11; DB 16; Length 492;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 FPNQWVDRIPT 233
DB 224 FPNQWVDRIPT 234

RESULT 5
Q92N07
ID Q92N07 PRELIMINARY; PRT: 494 AA.
AC Q92N07;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Probable mannitol 2-dehydrogenase protein (EC 1.1.1.67).
GN MTLK OR R02440 OR SMC01501.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gouzie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsperger U.,
RA Renard C., Thebaud P., Vandenberg M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591790; CAC47019.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PROSITE; PS00974; MANNITOL_DHGENASE; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 494 AA; 53822 MW; 7533287D8EF58F24 CRC64;

Query Match          2.3%; Score 11; DB 16; Length 494;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 FPNQWVDRIPT 233
DB 224 FPNQWVDRIPT 234

RESULT 6
Q8X516
ID Q8X516 PRELIMINARY; PRT: 486 AA.
AC Q8X516;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannonate oxidoreductase (D-mannonate oxidoreductase).
GN UXUB OR Z5921 OR EC55282.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

```

```

RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005663; AAG59505.1;
DR EMBL; AP002569; BAB38705.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 486 AA; 53656 MW; C519DD28763A41DC CRC64;

Query Match          2.1%; Score 10; DB 16; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 KAFTVNSCDN 190
DB 178 KAFTVNSCDN 187

RESULT 7
Q911D6
ID Q911D6 PRELIMINARY; PRT: 491 AA.
AC Q911D6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Mannitol dehydrogenase.
GN MTLDR OR PA2342.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004660; AAG05730.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 491 AA; 54350 MW; 8729A9F7E9841C39 CRC64;

Query Match          2.1%; Score 10; DB 16; Length 491;

```

```

Best Local Similarity 100.0%; Pred. No. 0.32; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Qy 182 AFTVMSCDNL 191
Db 183 AFTVMSCDNL 192
|||||

RESULT 8
O08355 PRELIMINARY; PRT; 493 AA.
AC O08355;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Mannitol 2-dehydrogenase (EC 1.1.1.67) (MDH).
CN MTD.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND SEQUENCE OF 1-20.
RC STRAIN=DSM 50106;
RX MEDLINE=97236441; PubMed=9116029;
RA Bruenker P., Altenbuchner J., Kulbe K.D., Mattes R.;
RT "Cloning, nucleotide sequence and expression of a mannitol
dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in
Escherichia coli.";
RT Biochim. Biophys. Acta 1351:157-167(1997).
RL [2]
RN SEQUENCE FROM N.A.
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 50106;
RX MEDLINE=98121321; PubMed=9461423;
RA Bruenker P., Altenbuchner J., Mattes R.;
RT "Structure and function of the genes involved in mannitol, arabinol
and glucitol utilization from Pseudomonas fluorescens DSM50106.";
RL Gene 206:117-126(1998).
CC -!- FUNCTION: A MANNITOL DEHYDROGENASE WITH A BROAD SUBSTRATE
SPECIFICITY. SUBSTRATES INCLUDE MANNITOL, ARABITOL AND SORBITOL,
THESE ARE OXIDIZED TO GIVE THE CORRESPONDING KETO SUGARS. THIS
ENZYME WILL ALSO CATALYZE THE REDUCTION OF FRUCTOSE AND XYLULOSE.
CC -!- CATALYTIC ACTIVITY: D-MANNITOL + NAD(+) -> D-FRUCTOSE + NADH.
CC -!- SUBUNIT: MONOMER.
CC -!- INDUCTION: BY MANNITOL, ARABITOL AND SORBITOL. NOT INDUCED BY
FRUCTOSE OR XYLULOSE.
CC -!- MISCELLANEOUS: THE MTDL PROTEIN IS ENCODED BY THE MTL EFGKDVZ
OPERON. THIS OPERON ENCODES PROTEINS FOR THE UPTAKE AND
UTILIZATION OF MANNITOL, ARABITOL AND SORBITOL.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
DR EMBL; AF007800; AAC04472.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTDLHDHGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 29 40 NAD (BY SIMILARITY).
SQ SEQUENCE 493 AA: 54497 MW; 7C12DFA443CEA443 CRC64;

Query Match 2.1%; Score 10; DB 2; Length 493;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDNL 191
Db 183 AFTVMSCDNL 192
|||||

RESULT 9
Q8X506 PRELIMINARY; PRT; 486 AA.
AC Q8X506;
DT 01-MAR-2002 (TREMBlrel. 20, Created)

```

```

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative oxidoreductase.
GN YDFI OR Z2155 OR ECS2151.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posafai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Apodaca J., Davis N.W., Lim A., Dimalanta E.T., Potamoukis K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005350; AAG56220.1;
DR EMBL; AP002557; BAB35574.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTDLHDHGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 486 AA: 53581 MW; 1AFD234D83108D9E CRC64;

Query Match 1.9%; Score 9; DB 16; Length 486;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDNL 190
Db 176 AFTVMSCDNL 184
|||||

RESULT 10
Q8ZPI8 PRELIMINARY; PRT; 488 AA.
AC Q8ZPI8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative mannitol dehydrogenase.
GN YDFI OR STM1508.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";

```

```

RL Nature 413:852-856(2001).
DR EMBL; AE008765; AAL20427.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Hypothetical protein; complete proteome.
SQ SEQUENCE 488 AA; 53896 MW; 704BFB8C27F4407A6 CRC64;

Query Match 1.9%; Score 9; DB 16; Length 488;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AFTVMSCDN 190
Db 176 AFTVMSCDN 184

RESULT 11
Q82626 ID Q82626 PRELIMINARY; PRT; 488 AA.
AC Q82626;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative D-mannonate oxidoreductase (EC 1.1.1.57).
GN STY1553.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN STRAIN=CT18;
RP SEQUENCE FROM N.A.
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL627270; CAD01805.1; -.
DR InterPro; IPR000569; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 488 AA; 53922 MW; 18F64DE2ACCA7781 CRC64;

Query Match 1.9%; Score 9; DB 16; Length 488;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AFTVMSCDN 190
Db 176 AFTVMSCDN 184

RESULT 12
Q8UD98 ID Q8UD98 PRELIMINARY; PRT; 244 AA.
AC Q8UD98;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE ABC transporter, substrate binding protein.
GN ATU2229 OR AGR_C_4053.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

```

```

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1];
RP SEQUENCE FROM N.A. PubMed=11743193;
RX MEDLINE=21608550;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Krespan W., Perry M.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2];
RP SEQUENCE FROM N.A. PubMed=11743194;
RX MEDLINE=21608551;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qureshi B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Hummel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009172; AAL43218.1; -.
DR EMBL; AE008138; AAK87971.1; -.
KW Complete proteome.
SQ SEQUENCE 244 AA; 27305 MW; 459F7E15EA3FF3B CRC64;

Query Match 1.6%; Score 8; DB 16; Length 244;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TLKSLPAN 13
Db 183 TLKSLPAN 190

RESULT 13
P74566 ID P74566 PRELIMINARY; PRT; 253 AA.
AC P74566;
DT 01-FEB-1997 (TReMBLrel. 02, Created)
DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein sir0654.
GN SLR0654.
OS Synecochystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecochystis.
OX NCBI_TaxID=1148;
RN [1];
RP SEQUENCE FROM N.A. PubMed=8905231;
RX MEDLINE=97061201;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuoka A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecochystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90916; BAA18673.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 28428 MW; F836F5042295F03B CRC64;

```

```
Query Match 1.6%; Score 8; DB 16; Length 253;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 VIPTLKAP 345
DB 2 VIPTLKAP 9
|||||

RESULT 14
Q9W439 PRELIMINARY; PRT; 261 AA.
AC Q9W439;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE CG4660 protein.
GN CG4660.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003436; AAF46120.2;
DR FlyBase; FBgn0029839; CG4660.
DR InterPro; IPR000365; 4HbcoA_thiostrse.
DR Pfam; PF03061; 4HBT; 1.
SQ SEQUENCE 261 AA; 29232 MW; 5AB58960057E3363 CRC64;
```

```
Query Match 1.6%; Score 8; DB 5; Length 261;
Best Local Similarity 100.0%; Pred. No. 21;
```

```
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 PELAKWIE 218
DB 241 PELAKWIE 248
|||||

RESULT 15
O24792 PRELIMINARY; PRT; 264 AA.
AC O24792;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Outer surface protein B (Fragment).
GN OSPB.
OS Borrelia garinii.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=29519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N333;
RA Fukunaga M., Yabuki M., Nakao M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007099; BAA22341.1; -.
DR HSSP; P14013; 1OSP.
DR InterPro; IPR001809; Outrsurface.
DR Pfam; PF00820; Lipoprotein_1; 1.
DR PRINTS; PR00968; OUTRSURFACE.
DR ProDom; PD001127; Outrsurface; 1.
FT NON_TER 264 264
SQ SEQUENCE 264 AA; 28671 MW; 7405E3CEB928EF08 CRC64;

Query Match 1.6%; Score 8; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 252 DLPLVAED 259
DB 45 DLPLVAED 52
|||||

Search completed: May 8, 2003, 16:44:41
Job time : 99 secs
```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:25:06 ; Search time 80 Seconds
(without alignments)
807.832 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516
Sequence: 1 MITRETLKSLPANVOAPPD.....KVIVLRKIIREKGVKAAIPA 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 13250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2516	100.0	485	21 AAB23172	Gluconobacter oxyd
2	2143	85.2	485	20 AAY49913	Gluconobacter subo
3	802	31.9	503	22 AAG89872	C. glutamicum prote
4	802	31.9	503	22 AAB79410	Corynebacterium gl
5	419.5	16.7	534	22 AAU39813	Propionibacterium
6	365.5	14.5	544	23 ABB54978	Lactococcus lactis
7	312.5	12.4	788	22 ABG13869	Novel human diagno
8	221.5	8.8	384	22 AAU35002	Enterococcus faeca
9	205	8.1	382	22 AAU34792	E. coli cellular p
10	203	8.1	382	13 AAR28826	Mannitol-1 dehydro

11	202.5	8.0	745	22	ABG28573	Novel human diagno
12	200	7.9	378	22	AAU38095	Streptococcus pneu
13	191.5	7.6	384	23	ABB53324	Lactococcus lactis
14	190	7.6	366	23	ABB05610	B. subtilis mannit
15	181	7.2	368	22	AAU34140	Staphylococcus aur
16	181	7.2	368	22	AAU36800	Staphylococcus aur
17	162	6.4	391	19	AAW37992	Mutant Aspergillus
18	162	6.4	391	20	AAW39873	A. oryzae DEBY932
19	162	6.4	391	23	ABB05608	Mutant Aspergillus
20	133.5	5.3	103	23	ABP33539	Human reductase-li
21	125.5	5.0	134	18	AAW28071	Mannitol-1-phospha
22	125	5.0	1504	22	AAW93729	Human protein sequ
23	122.5	4.9	1550	22	ABG20180	Novel human diagno
24	119	4.7	3685	10	AAW90373	Sequence encoded b
25	114.5	4.6	894	22	ABB59340	Drosophila melanog
26	114	4.5	183	22	ABG25111	Novel human diagno
27	110.5	4.4	7201	22	ABB71136	Drosophila melanog
28	109.5	4.4	2326	22	AAW50652	C. elegans spectri
29	106	4.2	3685	10	AAW90290	Human Duchenne mus
30	105.5	4.2	1560	21	AAW67203	Narbonolide synthase
31	105.5	4.2	1561	21	AAW77202	S. venezuelae pik
32	105.5	4.2	1562	21	AAW18639	Amino acid sequenc
33	105.5	4.2	1562	21	AAW77194	S. venezuelae macr
34	105.5	4.2	12199	21	AAW77180	S. venezuelae pik
35	105	4.2	1704	16	AAW70188	Arg-gingipain-2 pr
36	105	4.2	1704	18	AAW34843	Arg-gingipain high
37	105	4.2	1704	21	AAW67396	Arg-gingipain-2 am
38	105	4.2	1704	22	AAU08938	P. gingivalis high
39	104	4.1	588	17	AAW96209	Invasin-maltose bi
40	104	4.1	819	22	ABG20721	Novel human diagno
41	104	4.1	1867	23	ABG66751	Human novel polype
42	104	4.1	2059	23	ABP25711	Streptococcus poly
43	103.5	4.1	1533	22	ABG20179	Novel human diagno
44	103	4.1	395	22	AAW96802	Putative P. abyss
45	103	4.1	877	23	ABB55496	Lactococcus lactis

ALIGNMENTS

RESULT 1
AAB23172
ID AAB23172 standard; Protein; 485 AA.
XX
AC AAB23172;
XX
DT 29-JAN-2001 (first entry)
XX
DE Gluconobacter oxydans D-sorbitol dehydrogenase (SLDH).
XX
KW D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;
KW L-ascorbic acid biosynthesis; vitamin C.
XX
OS Gluconobacter oxydans.
XX
PN WO200055329-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-JP01608.
XX
PR 17-MAR-1999; 99JP-0072810.
XX
PR 06-AUG-1999; 99JP-0224679.
XX
PA (FUJI) FUJISAWA PHARM CO LTD.
XX
PI Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
PI Yamashita M, Takata Y;
XX
DR WPI: 2000-587530/55.
DR N-PSDB: AAA97430.
XX

PT Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
 PT culturing its gene-transformed host cells, useful for producing
 PT L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 acid production -

XX
 PS Claim 5; Page 58-60; 72pp; Japanese.

XX The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
 CC (SLDH; AAB23172) and to the gene encoding it (AA97430). SLDH has a
 CC molecular weight of about 54 kDa and catalyses the conversion of
 CC D-sorbitol into L-sorbose using NADP+ as a co-enzyme. SLDH specifically
 CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
 CC act on xylitol, ribitol, inositol and glycerol. The invention also
 CC encompasses expression vectors and host cells comprising the
 CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
 CC The invention further relates to a method for preparing L-sorbose by
 CC contacting the recombinant SLDH with D-sorbitol; a process for producing
 CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
 CC dehydrogenase and/or sorbose dehydrogenase with L-sorbose; and a
 CC process for preparing L-ascorbic acid or its alkaline earth metals salts
 CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
 CC L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 CC acid production. The present sequence represents the Gluconobacter
 CC oxydans SLDH protein.

XX
 SQ Sequence 485 AA;

Query Match 100.0%; Score 2516; DB 21; Length 485;

Best Local Similarity 100.0%; Pred. No. 3.9e-221;
 Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60

DB 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60

QY 61 GLTGSDRSKKKAEFEKQADCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120

DB 61 GLTGSDRSKKKAEFEKQADCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120

QY 121 PAIRIVSMITTEGGYININETTGAFDLENAAVKADLNKPEKPSVFGYVVEALRRWDAGG 180

DB 121 PAIRIVSMITTEGGYININETTGAFDLENAAVKADLNKPEKPSVFGYVVEALRRWDAGG 180

QY 181 KFTVMSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240

DB 181 KFTVMSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240

QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWWEYKIRMLNA 300

DB 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWWEYKIRMLNA 300

QY 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360

DB 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360

QY 361 FSNKMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRDLSRIAFGIASYLEMLRGRDEKG 420

DB 361 FSNKMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRDLSRIAFGIASYLEMLRGRDEKG 420

QY 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480

DB 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480

QY 481 AAIPA 485

DB 481 AAIPA 485

RESULT 2

AAV49913

ID AAY49913 standard; Protein; 485 AA.

XX

AC AAY49913;

XX 27-JAN-2000 (first entry)

XX Gluconobacter suboxydans L-sorbose reductase protein sequence.

XX Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
 mutant.

XX Gluconobacter suboxydans.

XX AU9920390-A.

XX 23-SEP-1999.

XX 11-MAR-1999; 99AU-0020390.

XX 13-MAR-1998; 98EP-0104546.

XX (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX Hoshino T, Tazoe M, Shinjoh M, Kon T;

XX WPI; 1999-579276/49.

XX N-PSDB; AAZ35672.

XX Set of genetically modified mutants not containing L sorbose reductase

XX Claim 8; Page 18-21; 33pp; English.

XX The present invention describes a genetically engineered
 CC microorganism derived from a microorganism belonging to the genus
 CC Gluconobacter or Acetobacter which is characterised in that the
 CC biological activity for reducing L-sorbose is substantially
 CC nullified by gene recombination gene. The present sequence represents
 CC Gluconobacter suboxydans L-sorbose reductase.

XX N.B. This patent is equivalent to the basic NO9901197 in week 199949.

XX Sequence 485 AA;

Query Match 85.2%; Score 2143; DB 20; Length 485;

Best Local Similarity 84.5%; Pred. No. 5e-187;
 Matches 410; Conservative 30; Mismatches 45; Indels 0; Gaps 0;

QY 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60

DB 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60

QY 61 GLTGSDRSKKKAEFEKQADCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120

DB 61 GLTGSDRSKKKAEFEKQADCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120

QY 121 PAIRIVSMITTEGGYININETTGAFDLENAAVKADLNKPEKPSVFGYVVEALRRWDAGG 180

DB 121 PAIRIVSMITTEGGYININETTGAFDLENAAVKADLNKPEKPSVFGYVVEALRRWDAGG 180

QY 181 KFTVMSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240

DB 181 KFTVMSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240

QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWWEYKIRMLNA 300

DB 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWWEYKIRMLNA 300

QY 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360

DB 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360

QY 361 FSNKMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRDLSRIAFGIASYLEMLRGRDEKG 420

DB 361 FSNKMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRDLSRIAFGIASYLEMLRGRDEKG 420

QY 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480

DB 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480

Db 421 GTYEPEPTFGDNHKTAKADDFESALKLPAPFADWNRDLTSGLNKKVVELKIIREKGVK 480

QY 481 AALPA 485

Db 481 AALPA 485

RESULT 3

ID AAG89872

AC AAG89872;

DT 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 3626.

DE C glutamicum protein fragment SEQ ID NO: 3626.

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS EP1108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

PF 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

PI WPI; 2001-376931/40.

DR N-PSDB; AAH65091.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

PT

PS Claim 17; SEQ ID NO: 3626; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX

SQ Sequence 503 AA;

Query Match 31.9%; Score 802; DB 22; Length 503;

Best Local Similarity 39.0%; Pred No. 2.3e-64;

Matches 168; Conservative 82; Mismatches 163; Indels 18; Gaps 6;

QY 2 ITRFTKSL--PANYQAPPYDIDGKPGIVHFGVGNFRAHEAFYVEQILEH--APDWAI 57

Db 7 LNTENLQETASTSGVQIPAFNADVPAGIVHFGVGFHRAHQAMYNELMNSGKALDWGI 66

QY 58 VGVGLTGSDRSKKAEFEKQAOCLSLTETAPSGKSTVVRMGALRDYLLAPADPEAVLKH 117

Db 67 ICMGVNPSDVRMRDA--LASQDHLTYLTTKAPDGTLDQKIIGSIIDYVFAPEDPARAVAT 124

QY 118 LVDPAIRIVSMITTEGGYNINETTGAFDLENAAVRADLK-----NPEKPTSTVFGYVVEAL 172

Db 125 LAQDSIRIVSLVTEGGYNIDPATDFDHTNPRIVADREALQAGDSTLTQTFEGLITAAL 184

QY 173 RRRWDAGGKAFVMSCDNLRHNGNVARKAFGLYAKARDELPALAKWIEENATFPNGMVDRI 232

Db 185 ISRKESGSTPFTIMSCDNIQNGDLAKRFFLAFAHVSSELGEMVNNVAFNPMVDRI 244

QY 233 PTVS--AEIAKLNAAAGLDDDLPLVAEDFHQWLEDOFADGRPLEKAGVQMGVDVTD 289

Db 245 PETTGDRDDIKEI---GYIDAPVVSSEDTQWLEDAFTQGRPAYEEVGVVSDVEP 300

QY 290 WEYVKIRMLNAGHVMVLCFPGILVGVENVDDAIEDSELLGNLKNLKNKDVIPTLKAPSGMT 349

Db 301 YELMKLRLLNASHQGLCYFGLHAGHHMVDVADTFQDFLLAYMEREATPTLKELPGVD 360

QY 350 LEGYRDSVISRFSNKAMSQDTRLIASDGCCKVQVFWTETVRRRAIEDKRLSRIAFGIAS 409

Db 361 LDAYRRQLIARFGNAVKDTVPRLCAESSDRIPKWLPPVVRNLAAGRDVTLSSAIVASW 420

QY 410 LEMLRGRDEKG 420

Db 421 ARYAEGTDEQG 431

RESULT 4

ID AAB79410

XX AAB79410 standard; Protein; 503 AA.

AC AAB79410;

DT 30-APR-2001 (first entry)

XX Corynebacterium glutamicum SMP protein sequence SEQ ID NO:336.

DE Corynebacterium glutamicum; carbon metabolism and energy production;

KW SMP protein; sugar metabolism and oxidative phosphorylation protein;

KW fine chemical production; organic acid; proteinogenic amino acid;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;

KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;

KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;

KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX

OS Corynebacterium glutamicum.

XX WO200100844-A2.

PN 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00943.

XX 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031431.

PR 08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.

PR 08-JUL-1999; 99DE-1031634.

PR 09-JUL-1999; 99DE-1032180.

PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032230.

PR 09-JUL-1999; 99US-0143208.

PR 14-JUL-1999; 99DE-1032924.

PR 14-JUL-1999; 99DE-1032973.

PR 14-JUL-1999; 99DE-1033005.

Query Match 14.5%; Score 365.5; DB 23; Length 544;
Best Local Similarity 26.4%; Pred. NO. 2.1e-24;
Matches 132; Conservative 84; Mismatches 213; Indels 71; Gaps 16;

AA 30-MAR-2001; 2001WO-US08631.
PF

Db 414 KIPRYGVTIGHYIANPRFSVKELEFIPLVIAAWCRYLIGINDELESFSPDPL----- 468

AA 30-MAR-2001; 2001WO-US08631.
PF

```
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS78056.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID No 44228; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AB00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 788 AA;
SQ
Query Match 12.4%; Score 312.5; DB 22; Length 788;
Best Local Similarity 37.6%; Pred. No. 2.6e-19;
Matches 67; Conservative 37; Mismatches 69; Indels 5; Gaps 3;
QY 255 LVAEDFHOMVLEDOFADGRPPLEKAGVGVGVDTWVEYVKIRMLNAGHVMCLFPGTLVGY 314
DB 3 VACEPRQWVIEDNFVAGRPWEKAGAEVLSDVLPTEEMKRLMNGSHSLAYLGLYLAGY 62
QY 315 ENVDADIEDSELGNLKNYLNDKVIPTLKAPSOITLEGYRDSYISRFSNKAMSDQTLRIA 374
DB 63 OHINDCMEDEHYRYAAYGLMLQEQAPTLAV-QGVDLQDYANRLIARYSNPALRHRTWQIA 121
QY 375 SDGCKVQVFWETVR--RAIEDKRLDSRIAGFIASYLEMLGRDRDKGTYESSEPT 430
DB 122 MDGSKRLPQRMLDVSRWHLAHDSKFDL--LALGVAGWMRYGVGVGDPGNIETSDPLF 177
RESULT 8
AAU35002
ID AAU35002 standard; Protein; 384 AA.
XX
AC AAU35002;
XX
DT 14-FEB-2002 (first entry)
DE Enterococcus faecalis cellular proliferation protein #289.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Enterococcus faecalis.
```

```
XX WO200170955-A2.
XX 27-SEP-2001.
XX 21-MAR-2001; 2001WO-US09180.
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX N-PSDB; AAS52861.
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10595; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 384 AA;
SQ
Query Match 8.8%; Score 221.5; DB 22; Length 384;
Best Local Similarity 22.6%; Pred. No. 1.8e-11;
Matches 83; Conservative 61; Mismatches 123; Indels 101; Gaps 14;
QY 29 VIFGVGNFPAHEAFYVEQILEHAPDWAIVGLTGSDRSKKAEFEKADCLYSLTETA 88
DB 4 VHFAGNIGRG---FGEIL-----AKNGFHITFVD-----VNET- 35
QY 89 PSGKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITITEGYNINETGAFDLEN 148
DB 36 -----IIQALKERKSYTIELADAS-----HQQINVENT-----GLNNTEPEKVVE 77
QY 149 AAVKADLKNPEKPSVFGVYVVEALRRRWDAAGKA-----FTVMSCDNLNRHNGNVARKAF 203
DB 78 ATAEDLVTTAIGPNILPRIAELIAQIGIDARAEANCQKPLDIIACENM-----IGGSTFL 132
QY 204 GYAKARDEPLAKWIEENA-----TFPNGWVDRIPTVSAAETAKKLNAAAGDLDPLVA 257
DB 133 A-----EEVAKYLNPAAYAEQWIGFPDAADVIRIVPLQKHE-----DPLFVGQV 174
QY 258 EDFHOWVLEDOFADGRPPLEKAGVGVGVDTWVEYVKIRMLNAGHVMCLFPGTLVGYENV 317
DB 175 EPFCEWVIDDTRKAK-EIQLEGVHVADLEPYIERKLFVSNTGATVATYTGALLGYQTI 233
```



```
PS Claim 22; Page 47 + 36-37; 55pp; English.
XX
CC Two separate genes are capable of inducing novel polyol biosynthesis
CC in transgenic plants. One gene is bacterial in origin, e.g. E. coli
CC derived mtd encoding mannitol-1-p dehydrogenase (AAQ31199). The gene
CC for this enzyme has been described by Lee and Salter, J. Bact., 153:2,
CC 685-692 (1983). The other gene is hml encoding myo-inositol O-
CC methyl transferase, derived from a stress tolerant plant (AAQ31200).
XX
SQ Sequence 382 AA;

Query Match      8.1%; Score 203; DB 13; Length 382;
Best Local Similarity 21.3%; Pred. No. 8.9e-10;
Matches 101; Conservative 75; Mismatches 160; Indels 138; Gaps 20;

QY 29 VHFVGNGFFRAEAYVEQILEHAPDWAIVGVGLTGSDBSKKAEFFKAQDC--LYSLTE 86
DB 4 LHFAGAGTGRG-----FIGLLADA-----GIQLTFADVQNVVLDAALNARHSYQHVHVE 53
QY 87 T----APSGKSTVRVMGALRDYLLAPADPEAYVLKHLVDPAIRIVSMITTEGGYNINETT 142
DB 54 TEQVDVFGVNAVSSIGDDVDVLIQVD--LVTNRVGPV----- 91
QY 143 AFDLENAVNAVKADLNKPEKSTVFGIVVEALRRRWAGGKAFVMSCDNLRHNGNVARKAF 202
DB 92 ----LERIA-----PAIAKGQV---KRKEQGNESPLNIIACENN--VRGTTQLK-- 131
QY 203 LGYAKARDPELAK-WTEENATPENGVDRTTPTVSAEIAKLNAAAGLDDDLPLVAEDPH 261
DB 132 -GHVNALPEDAKAWVEEHVGFVDSAVDRIVP-----SASATNDPLEVTVETFS 180
QY 262 QWVLE--DQFADGRPPLEKAGVQMGVDVTWVEYVKIRMLNAGHVMLCFPGILVGVENVDDA 320
DB 181 EWIVDKTEFGALPNI--PGMELTDNLMAFVERKLTTLNTGHAITAYLGLKLAGHOTIRDA 238
QY 321 IDESELLGNLKNYLNKDVPTTLKAPSGWTL-----EGYRDSVISRFSNKAMSDQ 369
DB 239 ILDEKIRAVVKGAMEB-----SGAVLIKRYGFDADKHAAYIQKILGRFENPYLKDD 289
QY 370 TLRASDGCSSKVQVFWTETVRRRAIEDKRLSRIAFGIAASYLEMLRGRDEKGYTESSEPT 429
DB 290 VERVG-----RQPLRLKSAGDRLIKPL-----GTLEWGLPH 321
QY 430 YGDAEWKLAADFFESSLKLPADFGWRDLDTSELQKVIVLRIKRIREKGVKAAI 483
DB 322 KNRIE-GIAAAMHFR-----SEDDPQAQELAALIAADKGPQAAL 358

RESULT 11
ABG28573
XX ABG28573 standard; Protein; 745 AA.
XX
XX
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #28564.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
```

```
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS92760.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
XX Claim 20; SEQ ID NO 58932; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 745 AA;

Query Match      8.0%; Score 202.5; DB 22; Length 745;
Best Local Similarity 30.8%; Pred. No. 2.8e-09;
Matches 45; Conservative 30; Mismatches 70; Indels 1; Gaps 1;

QY 283 MYGDVTDVEYKIRMLNAGHVMLCFPGILVGVENVDDATEDSELLGNLKNYLNKDVIP 342
DB 1 MYNDVLPWEEMKRLMLNGSHSLAYLGLYLSGFAHISDCMQDRAFRHARTLMDQAP 60
QY 343 KAPSGMTLEGYRDSVISRFSNKAMSDQTLRIASDGCSSKVQVFWTETVRRRAIEDKRLSRI 402
DB 61 QI-KVDVLTQYADKLIAREFANPALKHKTQWIAMDGSKLPQRLMAGIRITHOGRETDWSLL 119
QY 403 AFGIASYLEMLRGRDEKGYTESSEP 428
DB 120 ALGVAGWMRYVSGVDDAGNAIDVRDP 145

RESULT 12
AAU38095
XX AAU38095 standard; Protein; 378 AA.
XX
XX
XX AAU38095;
XX
XX 14-FEB-2002 (first entry)
XX
XX Streptococcus pneumoniae cellular proliferation protein #524.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Streptococcus pneumoniae.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
```


Db 137 VSDYSESKLYSKFI-----GPPAAVDRIYP-----AQHKDLYLYVEVEPF 180
Qy 261 HOWVLEQFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNAGHVMLCFPGILVGYENVDDA 320.
Db 181 SEMVIDASHLKKN-EIKLEGVHTTDLPEFIERKUFVSNGHAAVAYSAYGYKTILEG 239
Qy 321 IEDSELLGNLK-----NYLNKDVITPLKAPSGMTLEGYRDSVISRFSNKA 365
Db 240 LQDEEILNLIKAVOKETRALLLAKWAQYKQD-----ELIKYHELLIISRFSNPE 288
Qy 366 MSDQTLRIASDGCQVQVFWETVTRRAIDKRD 398
Db 289 IIDEVSRVARTPIRKLG--YDERFIRPIRELND 319

RESULT 14
ID ABB05610 standard; Protein; 366 AA.
AC ABB05610;
XX
DT 24-APR-2002 (first entry)
XX
DE B. subtilis mannitol-1-phosphate dehydrogenase protein SEQ ID NO:28.
XX
KW Aspergillus oryzae; mutant; modified; reduced transcription; hormone;
KW reduced translation; reduced secretion; receptor; antibody; reporter;
KW enzyme; lipase; mannitol-1-phosphate dehydrogenase.
XX
OS Bacillus subtilis.
OS Synthetic.
XX
XX US6323002-B1.
XX
PD 27-NOV-2001.
XX
XX 25-JUN-1999; 99US-0339972.
XX
PR 12-SEP-1997; 97US-0928692.
PR 13-SEP-1996; 96US-0713312.
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
XX
XX Brody H, Yaver DS, Lamsa M, Hansen K;
XX
DR WPI; 2002-163017/21.
XX
PT Producing a polypeptide using a cell for reducing the production of the
PT polypeptide, comprises inserting DNA into the genome of the cell at a
PT position not within the polypeptide coding sequence or a regulatory
PT sequence
XX
PS Example 15; Column 119-122; 129pp; English.
XX
CC The present invention describes a method for producing a polypeptide (P1)
CC comprising cultivating a mutant cell whose parent cell comprises a DNA
CC sequence encoding P1, by introducing a nucleic acid construct into the
CC genome of the parent cell at a locus not within the P1 sequence, so
CC that P1 transcription, translation or secretion is reduced, and
CC recovering P1. The method is used to produce a polypeptide, such as a
CC recombinant or heterologous hormone, hormone variant, receptor, antibody,
CC reporter or enzyme, particularly an oxidoreductase, transferase,
CC hydrolase, lyase, isomerase or ligase. The present sequence represents
CC a Bacillus subtilis mannitol-1-phosphate dehydrogenase protein which has
CC 34.7% identity with mutant Aspergillus oryzae DEBY932, which is used in
CC an example from the present invention.
XX
SQ Sequence 366 AA;

Query Match 7.6%; Score 190; DB 23; Length 366;
Best Local Similarity 22.5%; Pred. No. 1.3e-08;
Matches 67; Conservative 50; Mismatches 121; Indels 60; Gaps 10;

Qy 29 VHFVGYNFFRAHEAFYVQILEHAPDWAIVGVLGTSRDKKKAEEFKAOQCLYSLTETA 88
Db 4 LHFAGNIGRG---FIGALLHHS-----GYDVVF-----A 30
Qy 89 PSGKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITTEGGYINNETTGAFDLEN 148
Db 31 DYNETMVSLLNEKKEYTVELAE-EGRSSEIIGPVSAINSQSQTIELYRL-----MNE 81
Qy 149 AAVKADLANPEKPSVFCYVVEALRRRRDAGKAFVMSCDNLRHNGNVARKAFLYAKA 208
Db 82 AALITTAVGPNVNLKIAPSIAGLRRNTA--NTLNIACENMIGGSSFLKKEI--YSHL 137
Qy 209 RDELAKWIEENATPPNGMVDRIPTVSAETAKKLNAAAGLDDDLPLVAEDFHOWVLEDO 268
Db 138 TEAE-QKSVSETLGFNSAVDRIVPIQHE-----DPLKVSVEFFHVIDES 184
Qy 269 FADGRPPLEKAGVQMGVDVTDWEYVKIRMLNAGHVMLCFPGILVGYENVDDAIEDSEL 326
Db 185 GFKGKTPVIN-GALFVDDLTPTVIERKLTFTVNTGHAVTAYVGVQRLKTVKEAIDHPEI 241

RESULT 15
ID AAU34140 standard; Protein; 368 AA.
AC AAU34140;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #416.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
OS WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-20727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX
DR N-PSDB; AAS51999.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
PS
XX Example 3; Seq ID No 5636; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: the sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 368 AA;
Query Match 7.2%; Score 181; DB 22; Length 368;
Best Local Similarity 21.4%; Pred. No. 8.6e-08;
Matches 99; Conservative 64; Mismatches 168; Indels 132; Gaps 17;
QY 29 VHFVGNFFRAEAYVEQILEHADWAIVGVLGTGSDRSKKAEFKAQDCLYSLTETA 88
DB 4 VHFAGNIGRG----FIGYIL-----ADNNYKVTFADVNEEIIAAL-AHDHQYDVI-LA 51
QY 89 PSKSTVRVMGALRDYLLAPADPEAVLKHLDVDPAIRIVSMTITTEGGYNINETTGAFDLEN 148
DB 52 DESKTTTRV-----NNVDA-----INSMQPSEALKQ 77
QY 149 AAVKADLNKPEKSTVFGYVVEALRRRRWDAGGKFTVMSCDNLRHNGNVARKAFLGYAKA 208
DB 78 AILEADIITTAGVNIPLIIAKSFAPFLKEKTNHVNIVACENAIMATDLKKAVL----- 132
QY 209 RDELAKEEENATFNGMWDRIPTVSAEIAKLNAAAGLDDDLPLVAEDFHQWLEDO 268
DB 133 ---DITGPLGHNHFANSADVRIPLQNE-----NILDVMVEPEYEWVEKD 177
QY 269 FADGRPLEKAGVQMGVDVTWVEYKIRMLNAGHVMLCPGILVGVENVDDAIEDSELIG 328
DB 178 ANYG-PELNH--IKYVDDLTPYIERKLLTVNTGHAYLAYAGFAKATVLDRAVKDSSIEA 234
QY 329 NLKNYLKNDVITPLKAPSGMTLE-----GYRDSVISRFSNKAMSDQTLRIASDGCS 379
DB 235 GLRRVL-----AETSQVITNEFDTEAQAGYVEKIIDRENNSVLSDEVTRVGRG--- 284
QY 380 KQVQVFWETVRRRAIEDKRDLRSIAFGIASYLEMLRGD-EKGGTYESESSEPTYGDAEWKLA 438
DB 285 -----TLRKIGPKDR-----IKPLKLYLNKDLERTGLLNTAA-----LLL 320
QY 439 KADDFESSLKLPAFDGWRDLDTSELDOKVIVLKRITIREKGVA 481
DB 321 KYDD-----TADQETVEKNYIKEHGLKA 344

Search completed: May 8, 2003, 16:32:26
Job time : 90 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:22:01 ; Search time 29 Seconds
(without alignments)
492.073 Million cell

Title: US-09-926-163B-2
Perfect score: 2516
Sequence: 1 MITRETLSLPANVQAPPYD.....KVIVLRKIIIREKGVKAALPA 485

Scoring table: BLOSUM62

scoring table: BLOS0M02
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

FOSE PROCESSING. MINIMUM MATCH 0%
MAXIMUM MATCH 100%

Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

```

database :
issued_patents_Ar :
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*

```

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pe

6: /cgn2_6/ptodata/1/iaa/backfiles1.pe

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
1	202	8.0		382	1	US-08-186-833-2	Sequence 2, Appli
2	190	7.6		366	2	US-08-928-692-28	Sequence 28, Appli
3	190	7.6		366	4	US-09-339-972-28	Sequence 28, Appli
4	162	6.4	391	2	US-08-928-692-26	Sequence 26, Appli	Sequence 26, Appli
5	162	6.4	391	4	US-09-339-972-26	Sequence 26, Appli	Sequence 3, Appli
6	105.5	4.2	1562	3	US-09-320-878-3	Sequence 3, Appli	Sequence 35, Appli
7	105.5	4.2	1562	4	US-09-105-537-35	Sequence 6, Appli	Sequence 6, Appli
8	105.5	4.2	11877	4	US-09-108-537-6	Sequence 10, Appli	Sequence 10, Appli
9	105	4.2	1704	3	US-08-336-308A-10	Sequence 6, Appli	Sequence 10, Appli
10	105	4.2	1704	3	US-08-822-324-6	Sequence 10, Appli	Sequence 10, Appli
11	105	4.2	1704	4	US-09-490-931-10	Sequence 4, Appli	Sequence 4, Appli
12	104	4.1	588	5	PCT-US95-13749-4	Sequence 2, Appli	Sequence 2, Appli
13	102	4.1	407	4	US-08-955-957A-2	Sequence 2, Appli	Sequence 2, Appli
14	101	4.0	859	1	US-08-053-614-2	Sequence 2, Appli	Sequence 2, Appli
15	101	4.0	859	1	US-08-316-397B-2	Sequence 2, Appli	Sequence 2, Appli
16	101	4.0	859	2	US-09-034-306-2	Sequence 2, Appli	Sequence 2, Appli
17	101	4.0	859	4	US-09-259-437-2	Sequence 2, Appli	Sequence 2, Appli
18	101	4.0	859	5	PCT-US93-09782-2	Sequence 2, Appli	Sequence 2, Appli
19	101	4.0	1181	1	US-08-053-614-4	Sequence 4, Appli	Sequence 4, Appli
20	101	4.0	1181	1	US-08-316-397B-4	Sequence 4, Appli	Sequence 4, Appli
21	101	4.0	1181	2	US-09-034-306-4	Sequence 4, Appli	Sequence 4, Appli
22	101	4.0	1181	4	US-09-259-437-4	Sequence 4, Appli	Sequence 4, Appli
23	101	4.0	1181	5	PCT-US93-09782-4	Sequence 4, Appli	Sequence 4, Appli
24	97	3.9	1687	2	US-08-570-311-29	Sequence 2, Appli	Sequence 2, Appli
25	96.5	3.8	1181	2	US-08-488-940-2	Sequence 2, Appli	Sequence 2, Appli
26	95.5	3.8	570	1	US-08-403-866-6	Sequence 6, Appli	Sequence 6, Appli
27	95.5	3.8	821	1	US-09-377-465A-2	Sequence 2, Appli	Sequence 2, Appli

Sequence 48,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 3,	Appli
Sequence 2,	Appli
Sequence 3,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 2,	Appli
Sequence 23,	Appli
Sequence 23,	Appli

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 8.0%; Score 202; DB 1; Length 382;
Best Local Similarity 20.9%; Pred. No. 2.9e-11;
Matches 99; Conservative 70; Mismatches 168; Indels 1

QY 29 VHEGVGNFRAHEAFYVEQILEHAPDWAIVGVGLTGSDBSKKKAAEFKAQDCLYSLTETA 88

```
Db 4 LHFAGNIGRG-----FIGKLLADA-----GIQTFADVNVQVLDALNARH----- 44
QY 89 PSCKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITTEGGYINNETTGAFDLEN 148
Db 45 ---SYQVHVVGTEQ-----VD-----TVSGVNAVSGIGD-DVWD 75
QY 149 AAVKADLKPKPEKSTVGVVVEAL-----RRWDAGGKAFVTMSCDNLRHNGNVARAKFL 203
Db 76 LIAQVDLVTRVGVVVLRIAPAIKAGQVKKEQGNESPLNIIACENM-VRGTTQLK--- 131
QY 204 GYAKARDEPELAK-WIEENATFNGWVDRTPTVSAEIAKLNAAAGLDDDLPLVAEDPHQ 262
Db 132 GHVMNALPEDAKAWEHGVFVDSADVIRPP-----SASATNDPLEVTVEFSE 181
QY 263 WYLE-DOFADGPRPLEKAGQVQVGDVTDWEYVKIRMLNAGHVLMCFPGTLVGYENVDDAI 321
Db 182 WIVDKTFKALPNI--PGMELTDNLMAFVERKLTFTNGHTAITAYLGLAGHQIIRAI 239
QY 322 EDELLGNLKNYLKNDVIPTLKAPSGMTL-----EGYRDSVTSRFSNKAMSQOT 370
Db 240 LDEKIRAVVKGAMEE-----SGAVLIKRYGFDADKHAAYIQILGRFENPYLKDDV 290
QY 371 LRASDGCSKVQVFWTETVYRAIEDKRLSRIAFTAGIASVLEMLRGDEKGGTYESEPTY 430
Db 291 ERVG-----RQPLRKLISAGDRLIKPLL-----GTLEYGLPHK 322
QY 431 GDAEWKLAKADDFESSKLKLPADFGWRDLDTSELDOKVIVLRKIIREKGVKAAI 483
Db 323 NLIE-GIAAAMHFR-----SEDDPQAQELALIAADKGPQAL 358

RESULT 2
US-08-928-692-28
; Sequence 28, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727 of No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727e
```

```
US-08-928-692-28
Query Match 7.6%; Score 190; DB 2; Length 366;
Best Local Similarity 22.5%; Pred. No. 4e-10;
Matches 67; Conservative 50; Mismatches 121; Indels 60; Gaps 10;

QY 29 VHFVGNGFRAHEAFYVEQILEHAPDMAIVGVLGTGSDRSKKKAEFKAQCLXSLTETA 88
Db 4 LHFAGNIGRG-----FIGALLHHS-----GYDVF-----A 30
QY 89 PSCKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITTEGGYINNETTGAFDLEN 148
Db 31 DVNETMVSLLNEKKETVVELAE-EGRSSEITGVPVSAINSGSQTEELYRL-----MNE 81
QY 149 AAVKADLKPKPEKSTVGVVVEALRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 208
Db 82 AALITAVGNVVKLIAPSIAGELRRRTA--NTLIIACENMIGGSSFLKKEI--YSHL 137
QY 209 RDELAKWIEENATEPNGMVDRIPTVSAEIAKLNAAAGLDDDLPLVAEDPHQWVLEDO 268
Db 138 TEAE-QKSVSETLGFNSAVDRIVPIQHHE-----DPLKVSVEPFFEWVIDES 184
QY 269 FADGRPPLEKAGQVQVGDVTDWEYVKIRMLNAGHVLMCFPGTLVGYENVDDAIDSEL 326
Db 185 GFKGKTPVIN-GALEVDDLTPIERKLTFTVNTGHAVTAYGVQYRGLKTVKEAIDHPEI 241

RESULT 3
US-09-339-972-28
; Sequence 28, Application US/09339972
; Patent No. 6323002
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6323002 of No. 6323002disk of No. 6323002th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/339,972
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,692
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6323002e
; US-09-339-972-28
```

```
Query Match          7.6%; Score 190; DB 4; Length 366;
Best Local Similarity 22.5%; Pred. No. 4e-10;
Matches 67; Conservative 50; Mismatches 121; Indels 60; Gaps 10;

QY 29 VHFVGNNFFRAHEAFYVEQILEHAPDWAIVGVLGTSRDKKKAEEFKAQDCLYSLETETA 88
DB 4 LHFAGNIGRG---FVAGLLHHS-----GYDVVF-----A 30

QY 89 PSKSTVRVMGALRDYLLAPADPEAVLKHLDPAIRIVSMITTEGYNINETTGAFDLEN 148
DB 31 DVNETWVSLNKEKEYTVLAE-EGRSSIIIGPVSAINSGSTFEELYRL-----MNE 81

QY 149 AAVKADLNKPEKSTVFGYVVEALRRRWDAGGKFTVMSCDNLNRHNGNVARKAFGLCY 208
DB 82 AALITAVGPNVCLKIAPSAEGLRRRTA--NTLNIIACENNIGSSPLKKEI--YSHL 137

QY 209 RPELAKWIEENATPNGMVDRIPTVSAEIAKLNAAAGLDDDLPLVAEDFHQWVLEDO 268
DB 138 TEAE-QKSVSETLGPNSAVDRIVPIQHHE-----DPLKVSVEPEFVWIDES 184

QY 269 FADGRPPEKAGVQVGVDTWDEYVYKIRMLNAGHVLMCLFPGILVGVENVDDAIEDSEL 326
DB 185 GFKGKTPVIN-GALFVDDLTPTPIERKLFVTNTGHAVTAIVGQGLKTVKEAIDHPEI 241

RESULT 4
US-08-928-692-26
; Sequence 26, Application US/08928692
; Patent No. 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727o No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727e
US-08-928-692-26

Query Match          6.4%; Score 162; DB 2; Length 391;
Best Local Similarity 21.8%; Pred. No. 2.4e-07;
Matches 77; Conservative 55; Mismatches 149; Indels 72; Gaps 15;
```

```
QY 29 VHFVGNNFFRAHEAFYVEQILEHAPDWAIVGVLGTSRDKKKAEEFKAQDCLYSLETETA 88
DB 6 IOFGGNGIRG---FVAEFL-HAAGYEVFIDVMSVINSLQ-----QTPSYDVTVEVS 54

QY 89 PSKSTVRVMGALRDYLLAPADPEAVLKHLDPAIRIVSMITTEGYN-INETTGAFDLE 147
DB 55 EGEST-----KTIT--NYRAINSKTHEADV 79

QY 148 NAAVRADLNKPEKSTVFGYVVEALRRRWDG--GKFTVMSCDNLNRHNGNVARKAFGLCY 205
DB 80 QBIASADVTCVAGNLIKFIAPVIAKGIDARTERPVAIVACE---NAIGATDTLHGY 135

QY 206 AKAR-DPELAKWIEENATPNGMVDRIPTVSAEIAKLNAAAGLDDDLPLVAEDFHQWV 264
DB 136 IKOHTNPDRLETLSEARFANSADRIVPNQPPN-----SGLN---VRIKFIYEW 183

QY 265 LE-DOFAD-GRPPEKAGVQVGVDTWDEYVYKIRMLNAGHVLMCLFPGILVGVENVDDAIE 322
DB 184 VEKTPFGEHGHDI--PAIHWVDHLEPYIERKLFVTNTGHATTAYYAHKRGKKMIAEAL 241

QY 323 DSELLGNLKNYLNK--DVIPTLKAPSGMTLEGYRDSVISRFSNKAMSQDQLRI 373
DB 242 DPEIRETVHKVLEETASLIVSKHEISEQEKEYVDKIVSRISNPYLEDNVERV 294

RESULT 5
US-09-339-972-26
; Sequence 26, Application US/09339972
; Patent No. 6323002
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6323002o No. 6323002disk of No. 6323002th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/339,972
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,692
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6323002e
US-09-339-972-26
```

Query Match	6.48;	Score	162;	DB	4;	Length	391;
Best Local Similarity	21.88;	Prod.	No.	2.4e-07;			
Matches	77;	Conservative	55;	Mismatches	149;	Indels	72;
Gaps	15;						
Qy	29	VHFGVGNFRFAHEAFYVEQIILEHAPDWAIVGVGLTSDRSKKKAEFKAQDCLXSLTETA	88				
Db	6	IOFGGNGIRG-----FVAEFL-HAAGVEYFIDVMSVINSLQ-----QTPSDVTEVS	54				
Qy	89	PSGKSTVRVMGALRDYLLAPADPEAVLKLHLDVPAIRIVSMTITEGGYN-INETTGAFDLE	147				
Db	55	EEGEST-----KTIIT-NYRAINSKTHEADV 79					
Qy	148	NAAYKADLNKEPKSTVFEGYVVEALRRRWA--GGKAPTVMSCONLRINGNVAKAFILGY	205				
Db	80	OEIASADVTVCAVGNILKFTAPVIAKIDATERPVAVIAE---NAIGATDTLHG	135				
Qy	206	AKAR-DPELAKWIEENATFPNGMDVRITPTYSAEIAKLNAASGLDDDLPLVAEDFHQW	264				
Db	136	IKQHTNPDRLTTLSEARAFANSAIDRIVPNOPN-----SGLN-----VRIEKFYEWA	183				
Qy	265	LE-DQFAD-GRPPLKAGVQVGVDTWEYVKIRMLNAGHVMLCPFGPILGVENVDDAIE	322				
Db	184	VEKTPFGGWGHPDI--PAIHVVDHLEPIYERKLTFTVNGHTATTAYAHKRGCKMIAEAL	241				
Qy	323	DSELLGNLKNYLNK--DVIPITLKAPSGMTLGGYRDSVTSRFSNKRMSQOTURI	373				
Db	242	DPEIRETVHKVLEETASLIVSKHIEISEOEQKEYVDKIVSRISNPNVLENDVERV	294				

RESULT 6
 US-09-320-878-3
 ; Sequence 3, Application US/09320878A
 ; Patent No. 6117659
 ; GENERAL INFORMATION:
 ; APPLICANT: ASHLEY, Gary
 ; APPLICANT: BETLACH, Melanie C.
 ; APPLICANT: BETLACH, Mary C.
 ; APPLICANT: MCDANIEL, Robert
 ; APPLICANT: TANG, Li
 ; TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
 ; FILE REFERENCE: 300622002120
 ; CURRENT APPLICATION NUMBER: US/09/320,878A
 ; CURRENT FILING DATE: 1999-05-27
 ; EARLIER APPLICATION NUMBER: CIP OF 09/141,908
 ; EARLIER FILING DATE: 1998-08-28
 ; EARLIER APPLICATION NUMBER: CIP OF 09/073,538
 ; EARLIER FILING DATE: 1998-05-06
 ; EARLIER APPLICATION NUMBER: CIP OF 08/846,247
 ; EARLIER FILING DATE: 1997-04-30
 ; EARLIER APPLICATION NUMBER: 60/119,139
 ; EARLIER FILING DATE: 1999-02-08
 ; EARLIER APPLICATION NUMBER: 60/100,880
 ; EARLIER FILING DATE: 1998-09-22
 ; EARLIER APPLICATION NUMBER: 60/087,080
 ; EARLIER FILING DATE: 1998-05-28
 ; NUMBER OF SEQ ID NOS: 34
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 3
 ; LENGTH: 1562
 ; TYPE: prt
 ; ORGANISM: Streptomyces venezuelae
 US-09-320-878-3

	Query Match	4.28;	Score 105.5;	DB 3;	Length 1562;
	Best Local Similarity	20.8%;	Pred. No. 0.74;	Matches 100;	Conservative 74; Mismatches 194; Indels 113; Gaps 21.
QY	87	TAPSGKSTVRVYM-GALRDYL LAPDPEAVLKH----	LVDPAIRIVSMTIEGGVINNET	140 :	: :
Dd	313	TAPHGPSOQRIRLADARLARLPAGDVVEAHGTGTRUGDP-IEAQAIIATYG-----OEK	367	:	: :
QY	141	TGAFDLENAAYVKALKNPKPSTVFVGYY--YEALRR-----RWDAGSKA	182	: : : :	: :

```

368 SSEQPLRLGALKSNIGHTQAAAGVAGVYKVMQAMRHGLLPTLLHVDEPSDQIDWSAGTVE 427
183 FTVMSCD-NLRHNGNVARKAFLGYAKARDP-----ELAKWTEENATF--PNG-----MV 228
428 LLTEAVDWPEKQDGLRRAAVSSFGISGTAHVHVLLEAPAVEDSPAEPAGGGVVPWPV 487
229 DRIPT-TVSAIEAKLNAASGLDDDLPLVAEDFHQWVLEDFADGRPPLEKAGVOMGDV 287
488 SAKTPAALDAQIGLAAVADGRTVDPAVA-----ARALVDSRTAMEHRAV-AVGDS 538
288 TDWEVVKIRM-----LNAGHVMLCFPG-----ILVGYENVDDATEDSELGLNKNY 333
539 REALRDALRMEPEGLVGTSSDVRGVAFFPGQGTQWAGMGAEGLDSSPEFRAASMAECETA 598
334 LNKDVIPTLKA-----PSGMTLE-----GYRDSVISRFSNKAMSDQTL----- 371
599 LSRVYDMSLEAVVRQEPGAPTLDRVDVVQVPTFAVMVSLAKWQHGGITPQAVVGHSGOE 658
372 -----RIASDGCSSKVOVFWETVTRRAIEDKRDLSRIAGFIASYLEMLRGRDEKG--- 420
659 IAAAYVAGALTDDAARVVTLRKSSIAAHLAKGKGMISLALDEAAVLKRLSDFDGLSVAA 718
421 ----GTVESSEPYGDAEWKLAADDPESSILKLPADFGRWRDLDTSELDDQKIVILVKIIR 475
719 VNGPTATVWSGDPQTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEITEKELA 770
QY 476 E 476
DB 771 E 771

RESULT 7
US-09-105-537-35
; Sequence 35, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D. H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438U51
; CURRENT APPLICATION NUMBER: US/09/105.537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-35

Query Match 4.2%; Score 105.5; DB 4; Length 1562;
Best Local Similarity 20.8%; Pred. No. 0.74;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;

QY 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRVSTMTTEGGYNINET 140
DB 313 TAPGPGSQRVIRRALADARLAPGDVVVEAHGTGTRLGD-IEAQAALIATYG-----QEK 367
QY 141 TCAFDLENAAYKADLKNPKEPSTVFGYV--VEALRR-----RWDAGGA 182
DB 368 SSEQPLRLGALKSNIGHTQAAAGVAGVYKVMQAMRHGLLPTLLHVDEPSDQIDWSAGTVE 427
QY 193 FTVMSCD-NLRHNGNVARKAFLGYAKARDP-----ELAKWTEENATF--PNG-----MV 228
DB 428 LLTEAVDWPEKQDGLRRAAVSSFGISGTAHVHVLLEAPAVEDSPAEPAGGGVVPWPV 487
QY 229 DRIPT-TVSAIEAKLNAASGLDDDLPLVAEDFHQWVLEDFADGRPPLEKAGVOMGDV 287
DB 488 SAKTPAALDAQIGLAAVADGRTVDPAVA-----ARALVDSRTAMEHRAV-AVGDS 538
QY 288 TDWEVVKIRM-----LNAGHVMLCFPG-----ILVGYENVDDATEDSELGLNKNY 333

```


Db 539 REALDRLRMEGLVRGTSDVGRVAFVFPQGCTQWAGMGAELLDSSPEFAASMAECETA 598
QY 334 LNKDVIPTLKA-----PSGMTLE-----GYRDSVISRESNKAMSQDTL----- 371
Db 599 LSRVDMVSLVAVVRQEPGAPTLDRVDVQVPVFAVMVSLAKVWQHGGITPQAVVGHSGQE 658
QY 372 -----RIASDGCCKVQVFTETVRRRAIEDKRDLRSIAFGTASYLEMLRGRDEKG--- 420
Db 659 IAAAYVAGALTLDAAARVVTLRSKSTAAHLAGKGMISLALDEAAVLRKLSDFDGLSVAA 718
QY 421 -----GTYESSEPTYGDAEWKLAADDFESSLKLPAFDGWRDLDTSELQKQVIVLRKIIR 475
Db 719 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIEKELA 770
QY 476 E 476
Db 771 E 771
RESULT 8
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRM
; ORGANISM: Streptomyces venezuelae
US-09-105-537-6
Query Match 4.2%; Score 105.5; DB 4; Length 11877;
Best Local Similarity 20.8%; Pred. No. 20;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;
QY 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMITTEGGYNINET 140
Db 9001 TAPHGFSQQRVIRRALADARLAPGDVDVVEAHCTGTRLGD-LEAQLIATYG-----QEK 9055
QY 141 TGAFDLENAAVRADLKNPEKSTVFQYV--VEALRR-----RWDAGGKA 182
Db 9056 SSEQPLRLGALKSNIGHTQAAAAGVAGVIRKMQAMRHGLPKTLHVDEPSDQIDWSAGTVE 9115
QY 183 FTVMSCD-NLRHNGNVARKFLGYAKARDP-----ELAKWIEENATF--PNG-----NV 228
Db 9116 LITEAVDWEKQDGGURRAVSSFGISGTAHVLEAPAVEDSPAVEPPAGGVVPPVPV 9175
QY 229 DRITP-TVSAEIAKLNAAAGLDDLLPLVAEDFQHWLEQDQADGRPPLEKAGVQMVGDV 287
Db 9176 SAKTPAALDAQICOLAAAYADGRDVIDPVA-----ARALVDSSTAEHRAV-AVGDS 9226
QY 288 TWYEVYKIRM-----LNAGHVMCLCPFG-----ILVGYENVDDAIEDSELLGNLKNY 333
Db 9227 REALDRLRMEGLVRGTSDVGRVAFVFPQGCTQWAGMGAELLDSSPEFAASMAECETA 9286
QY 334 LNKDVIPTLKA-----PSGMTLE-----GYRDSVISRESNKAMSQDTL----- 371
Db 9287 LSRVDMVSLVAVVRQEPGAPTLDRVDVQVPVFAVMVSLAKVWQHGGITPQAVVGHSGQE 9346
QY 372 -----RIASDGCCKVQVFTETVRRRAIEDKRDLRSIAFGTASYLEMLRGRDEKG--- 420
Db 9347 IAAAYVAGALTLDAAARVVTLRSKSTAAHLAGKGMISLALDEAAVLRKLSDFDGLSVAA 9406

QY 421 -----GTYESSEPTYGDAEWKLAADDFESSLKLPAFDGWRDLDTSELQKQVIVLRKIIR 475
Db 9407 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIEKELA 9458
QY 476 E 476
Db 9459 E 9459
RESULT 9
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-308A-10
Query Match 4.2%; Score 105; DB 3; Length 1704;
Best Local Similarity 21.6%; Pred. No. 0.95;
Matches 99; Conservative 60; Mismatches 186; Indels 114; Gaps 22;
QY 86 ETAPSGKSTVRVMGALRDYLLAPADPEAVLKHLDVDPAIRIVSMITTEGGYNINETTGAFD 145
Db 752 QVPSDTHTLWPCNSVPANLFAPF--EYTVPENADPSCSPTNM-IMDGTASVINIPAGTYD 808
QY 146 LENAAYKADLK---NPEKPSVFGYVVEA-----LRREWDAG----- 179
Db 809 FAIAAPQANAKIWIAGQGPTKEDDYVEAGKKVHFLMKMGSGDGTFLTISEGGSDYTY 868
QY 180 -----CKAFTVMSCDNLRHNGNVARKFLGYAKARDPELAKWIEENATFPNGMV 228
Db 869 TVYRGTIKIEGLTATTFEEDGVA-TGNHECYCEVKYTAGVSPKVC KDVTVEGSNEFAPV 927


```
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-490-931-10

Query Match          4.2%; Score 105; DB 4; Length 1704;
Best Local Similarity 21.6%; Pred. No. 0.95;
Matches 99; Conservative 60; Mismatches 186; Indels 114; Gaps 22;

QY 86 ETAPSKSVRVNMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITTEGGYNINETTGAED 145
Db 752 QVTPTDTHLWPCNSVPANLFAFP--EYTVPENADSPSCPTNN-IMDGTASVNIPAGTYD 808
QY 146 LENAAYKADLK--NPEKSTVFYVVEA-----LRRRWADAG-----179.
Db 809 FATAAPQANAKIWIAGOGPTKEDDYVFEAGKYHFLMKKMGSDGTETLISEGGGSDITY 868
QY 180 -----GKFTVMSCDNLRHNGNVARKAFLOYAKARDELAKEWIEENATFPNGMV 228
Db 869 TVYRDGTKIKEGLTATTFEEDGVA-TGNEYCVVEVKYTAGVSPKCKDVTVEGSNEFAPV 927
QY 229 DRIT-PTVSAETAKLNASGLDDDLPLVAEDFHOWVLEQDADGRPPLEKAGVQVMGDV 287
Db 928 QNLTSAGVOKVTLKWDAPGNPNPNPNPTGTTLTSEFENGIPASWKT-IDADGDC 986
QY 288 TDWEYVKIRMLNAGHVMLCFPGILVGYENVDDAIEDSELLGNL-----KNYLNKDVIP 342
Db 987 HGWK-----PGNA-----PGI-AGYNSNGCVYSEFGLGGIGVLTDPNYL---ITPAL 1030
QY 343 KAPSGMTLEGYRDSVTSRFSNKAQSDOTLRIASDGCSKVQVFWTETVRRRAIEDKRLSRI 402
Db 1031 DLPNGGKLTFW---VCAQADANVASEHYAVVASTG-NDASNFTNALLETITAK----- 1080
QY 403 AFGIASYLEMLRCR-----DEKGGT-----YESSEPTYGDAEWKLAKA----- 440
Db 1081 --GVRS-PEAIRIGIQGTWRQKTVDPALPACTKYVAFHFQSTDMFYIDLDEVEIKANGKRA 1137
QY 441 ---DDFESSL--KLPA-----FDGWRDLDTSELD 464
Db 1138 DFTETESSTHGEAPEWTTIDADGCGQWLCCLSSQLD 1176

RESULT 12
PCT-US95-13749-4
; Sequence 4, Application PC/TUS9513749
; GENERAL INFORMATION:
; APPLICANT: Amgen Inc.
; TITLE OF INVENTION: COMPOSITIONS FOR INCREASED
; BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Denavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13749
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 588 amino acids
; TYPE: amino acid

; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US95-13749-4

Query Match          4.1%; Score 104; DB 5; Length 588;
Best Local Similarity 19.9%; Pred. No. 0.21;
Matches 105; Conservative 66; Mismatches 176; Indels 180; Gaps 29;

QY 17 PPVIDIDGKPGIVHFGVGNFFRAHEAF--YVQQ--ILEHAPQWAIIVGVLGTGSDRSKKKA 72
Db 74 PQVATGCDPDII-----FWAHDREFGYAOSGLLAETTPDKAF-----111
QY 73 EEFKAQDCILYSLTETAPSGKSTVRVNGALRDYLLA-----PADPE-----112
Db 112 -----QDKLYPETWDA-----VRYNGKLIAYPIAEVALLIYNKDLLPNPKTWEEIPA 160
QY 113 -----AVLKHLVDPAIRIVSMITTEGGYNINETTGAEDL-----ENRAVKADL- 155
Db 161 LKELKAKGKSALMFNLQEPYF-TWPLIAADGGYAFKYENGYDIKDVGVDNAGAKAGLT 219
QY 156 -----KNPEKSTVFYVVEALRRRWADAGGKAFV-----MSCDNLRHNGVA----- 198
Db 220 FLVDLILKKNHMADTDYSTAEA---AFNGETAMTINGPWANSNIDTSKVNYGTVLPTF 276
QY 199 ----RKAFGLYAKARDELAKEWIEENATFPNGMVDTRITPTVSAEIAKKL-----NAA 246
Db 277 KGQSPKPFVGVLSA-----GINAASPN-----KELAKEFLENYLLTDEGL 316
QY 247 SGLDDDLPLVAEDFHOWVLEQDFA-DGRPPLEKAGVQVMGDVTD-----WEYVKIRM 297
Db 317 EAVNKDKPLGVALKSY---EELAKDPRIATMENAQK-GEIMPNIPOHSAFWYAVRTAV 373
QY 298 LNAAGHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAPSGMTLEGYRDSV 357
Db 374 INAAS-----GRTVDEALKDAQ-----TRITK--VPTL---TGILVNGONFAT 412
QY 358 ISRFSNKAQSDOT--LRIASDGCSKVQVFWTETVRRRAIEDKRLSRIAFIASYLEM-LR 414
Db 413 DKGFPKTIKFNATFQLOMDNVANNNTQYEWSSSFENPV-SVNDGGOVTTITYYSEVAVT 471
QY 415 GRDEKGGTYESSEPTYGDAEWKLAKADDFESSLKLPAFGWRDLDT 461
Db 472 AKSKFPYSYSVSYRFPN-RW-----IYDGRSLVSS 502

RESULT 13
US-08-955-957A-2
; Sequence 2, Application US/08955957A
; Patent No. 6312920
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Rostock Jr., Paul R.
; TITLE OF INVENTION: SAM Operon
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955,957A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:24:41 ; Search time 53 Seconds
(without alignments)
842.122 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516

Sequence: 1 MITRETLKSLPANVOAPDY.....KVIVLRKTIKRGVKAIPA 485

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptoddata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptoddata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptoddata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptoddata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptoddata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptoddata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptoddata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptoddata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptoddata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptoddata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptoddata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptoddata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptoddata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptoddata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	802	31.9	503	9	US-09-738-626-3626
2	221.5	8.8	384	10	US-09-815-242-10595
3	205	8.1	382	10	US-09-815-242-10385
4	200	7.9	378	10	US-09-815-242-13688
5	181	7.2	368	10	US-09-815-242-5636
6	181	7.2	368	10	US-09-815-242-12393
7	125	5.0	1504	9	US-09-932-145-7
8	107	4.3	1703	9	US-09-824-574-3
9	107	4.3	1703	10	US-09-801-368-340
10	105.5	4.2	1562	9	US-09-860-846-35
11	105.5	4.2	1562	9	US-09-988-384B-35
12	105.5	4.2	1562	10	US-09-861-289-35
13	105.5	4.2	11877	9	US-09-860-846-6
14	105.5	4.2	11877	10	US-09-861-289-6
15	105.5	4.2	12199	9	US-09-988-384B-6
16	103	4.1	669	9	US-10-083-357-1336
17	102	4.1	407	9	US-09-943-702-2
18	100.5	4.0	1338	10	US-09-402-100-4
19	100	4.0	762	10	US-09-815-242-11058

20	99.5	4.0	2910	9	US-10-124-800-2	Sequence 2, Appli
21	98	3.9	559	10	US-09-815-242-11904	Sequence 11904, A
22	98	3.9	852	9	US-09-738-626-6399	Sequence 6999, Ap
23	97.5	3.9	736	9	US-09-978-295A-526	Sequence 526, App
24	97.5	3.9	736	9	US-09-978-697-526	Sequence 526, App
25	97.5	3.9	736	9	US-09-978-192A-526	Sequence 526, App
26	97.5	3.9	736	9	US-09-998-832A-526	Sequence 526, App
27	97.5	3.9	736	9	US-09-978-189-526	Sequence 526, App
28	97.5	3.9	736	9	US-10-174-590-420	Sequence 420, App
29	97.5	3.9	736	9	US-10-176-758-420	Sequence 420, App
30	97.5	3.9	736	9	US-10-175-737-420	Sequence 420, App
31	97.5	3.9	736	9	US-10-173-706-420	Sequence 420, App
32	97.5	3.9	736	9	US-10-175-738-420	Sequence 420, App
33	97.5	3.9	736	9	US-10-175-752-420	Sequence 420, App
34	97.5	3.9	736	9	US-10-176-482-420	Sequence 420, App
35	97.5	3.9	736	9	US-10-176-757-420	Sequence 420, App
36	97.5	3.9	736	9	US-10-176-913-420	Sequence 420, App
37	97.5	3.9	736	9	US-10-180-552-420	Sequence 420, App
38	97.5	3.9	736	9	US-10-180-557-420	Sequence 420, App
39	97.5	3.9	736	9	US-10-173-700-420	Sequence 420, App
40	97.5	3.9	736	9	US-10-174-572-420	Sequence 420, App
41	97.5	3.9	736	9	US-10-174-579-420	Sequence 420, App
42	97.5	3.9	736	9	US-10-174-582-420	Sequence 420, App
43	97.5	3.9	736	9	US-10-175-588-420	Sequence 420, App
44	97.5	3.9	736	9	US-10-175-739-420	Sequence 420, App
45	97.5	3.9	736	9	US-10-175-740-420	Sequence 420, App

ALIGNMENTS

RESULT 1

US-09-738-626-3626
Sequence 3626, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 3626
LENGTH: 503
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-3626

Query Match 31.9%; Score 802; DB 9; Length 503;

Best Local Similarity 39.0%; Pred. No. 7.4e-63;

Matches 168; Conservative 82; Mismatches 163; Indels 18; Gaps 6;

QY 2 ITRETLKSL--PANVOAPPYDIDGKPGVHVGFFRFAHEAFVVEQILEH--APDWAI 57

DB 7 LNTENLQELASTSGVQIEAFNADVPAGVHVGFGVGFRAHQAWYLNELMNECKALDWGI 66

QY 58 VGVGLTSGDRSKKAEFPKQDCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKH 117

```
Db 67 ICGVMPDSVRMRDA--LASQDHLTYLTAKPDGTLDOQKIIGSIIDYVFAPEDPARAVAT 124
Qy 118 LVDPATRIYVSMITTEGGYNINTEGAFDLENAAVKADLK-----NPKSPSTVFGYVWEAL 172
Db 125 LAQDSIRIVSLVTEGGYNIDPATEDFDITNPRIVADREALQAGDTSTLTQTFEGLITAL 184
Qy 173 RRRWDAGGKAFVMSCDNLNRHNGNVARKAFYAKARDPELAKWIEENATFPNGMVDRT 232
Db 185 ISKESGSTPTTMSCDNIQNGDLAKRFLAFASHVSSELGEWENNVAFPNSMVDRT 244
Qy 233 PTVS---AIAIKLNAASGLDDDLPLVAEDFHQWLEDFADGRPLEKAGVQMVGVDYD 289
Db 245 PETTDGDRDIKEI---GYIDAWPVSEDFQWLEDAFTQRPAYEEVGQVVSDEP 300
Qy 290 WEVYKRLNAGHVMLCPFGILVGVENVDDAJEDSELGNLKNLKNKDVIPTLKAPSGMT 349
Db 301 YELMKRLNLNASHQGLCYGHLGHVHWDVADRFQDFLLAYWEREATPTLKELPGVD 360
Qy 350 LEGYRDSVISRFSNKAASQTLRIASDGCYSKVQVFWTETVRRRAIEDKRLSRIAFGIASY 409
Db 361 LDAYRQRLIARFCGNAAVKDTVPRLCAESSDRIPKWLPPVRENLAAGRDVTLSSAIVASW 420
Qy 410 LEMLRGRDEKG 420
Db 421 ARYAEGTDEQG 431
```

RESULT 2

```
US-09-815-242-10595
; Sequence 10595, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10595
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10595
```

Query Match 8.8%; Score 221.5; DB 10; Length 384;
Best Local Similarity 22.6%; Pred. No. 1.9e-11;
Matches 83; Conservative 61; Mismatches 123; Indels 101; Gaps 14;

Qy 29 VHFVGNGNFRFAEAFVVEQILEHAPDWAIVGVLGTSDRSKKKAEEFKAQDCLYSLTETA 88

```
Db 4 VHFAGAGNTRG---FIGEIL-----AKNGFHITFVD-----VNET- 35
Qy 89 PSKSTVRVMGALRDYLLAPADPEAVLKLHVPAPTRIVSMITTEGGYNINTEGAFDLEN 148
Db 36 -----IIQALKERKSYTTELADAS-----HQQINVENVT-----GLNNWTEPEKVVE 77
Qy 149 AAVKADLKNPEKSPSTVFGYVVEALRRRWDAGKA-----FTVMSCDNLNRHNGNVARKAPL 203
Db 78 ATAEDADVTTAIGPNLPLRIELIAQIGIDARAEANCQKPLDIIACENM-----IGGSTFL 132
Qy 204 GYAKARDELAKWIEENA-----TFPNGMVDRTPTVSAEIAKLNAAASGLDDDLPLVA 257
Db 133 A-----EVAKYLNKPAYAEQWIGFPPDAADVRIVPLOKHE-----DFLFGVQV 174
Qy 258 EDFHFWLEDFADGRPLEKAGVQMVGVDVTDWEVYKRLNAGHVMLCPFGILVGYENV 317
Db 175 EFCFCEWIDDTNRKAK-ELQEGVHYVADLEPIYERKLFVSNTGHTATVAYTGALLGYQTI 233
Qy 318 DDAIEDSELLGNLKNLKNKDVIPTLKAPSGMTL-----EGYRDSVISRFSNKAAM 366
Db 234 DEAMQDALVVAQLKSVLOE-----TGKLLVAKWNFDEQEAHAAYTEKIIORFQNKYI 284
Qy 367 SDQTLRIA 374
Db 285 SDAITRVA 292
```

RESULT 3

```
US-09-815-242-10385
; Sequence 10385, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10385
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10385
```

Query Match 8.1%; Score 205; DB 10; Length 382;
Best Local Similarity 21.9%; Pred. No. 5.4e-10;
Matches 104; Conservative 71; Mismatches 161; Indels 138; Gaps 20;

Qy 29 VHFVGNGNFRFAEAFVVEQILEHAPDWAIVGVLGTSDRSKKKAEEFKAQDC--LYSLTE 86


```

Db 4 LHFAGNIGRG-----FIGLLADA-----GIQLTADVQVVLDAALNARHSQVHVGE 53
QY 87 T-----APSGKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMTITTEGGYNINETG 142
Db 54 TEQVDTVSGVNAVSSIGDDVDVLIQVD-----LVTTAV-----G 88
QY 143 AFOLENAVKAADLKNPEKSTVFEGYVEALRRWDAGGKAFVMSCDNLRHNGNVARKAF 202
Db 89 PVVLERIA-----PAIAGQV-----KRKEOGNESPLNIIACENM-VRGTTOLK-- 131
QY 203 LGVAKARDPELAK-WIEENATFNGMVDRIPTVSAEIAKLNAASGLDDDLPLVAEDPH 261
Db 132 -GHVMNALPEDAKAWVEEHVGFVDSAVDRIVP-----SASATNDPLEVTVETES 180
QY 262 OWYLE-DQFADGPPPLEKAGVQMGVDVTWEYVKIRMLNAGHVMCLFCPGLVLGVYENVDDA 320
Db 181 EWIVDTQFGALPNI--FGMELTDNLMAFVERKLTFTLTGHTAITAYLKLAGHQIIRDA 238
QY 321 IEDELLGNLKNYLNDVIPTLKAPSGMTL-----EGYRDSVISRFSNKAMSDQ 369
Db 239 ILDEKIRAVVKGAMEE-----SGAVLIKRYGFDADKHAAYIQILGRFENPVLKDD 289
QY 370 TIRIADGSKVQVFTETVRRRAIEDKRLSRTAFGIAASYLEMLGRDEKGGTYESSEPT 429
Db 290 VERVG-----RQPLRKLISAGDRLIKPLL-----GTLEYGLPH 321
QY 430 YGDAEWKLAKADDFESSLKLPAFGDWRDLTSELDOKVIVLRLKIIREKGVKAAI 483
Db 322 KNLIE-CIAAAMHFR-----SEDDPQAOELAAIADKGPQAAI 358

```

```

RESULT 4
US-09-815-242-13688
; Sequence 13688, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13688
; LENGTH: 378
; TYPE: PRN
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13688

```

Query Match 7.9%; Score 200; DB 10; Length 378;

```

Best Local Similarity 22.6%; Pred. No. 1.5e-09;
Matches 106; Conservative 71; Mismatches 163; Indels 130; Gaps 21;
QY 29 VHFQVGNFTRAHEAFVVEOIL-----EHPADWAIVGVLTCSDRSKKKAEFFKAQDCLYSL 84
Db 5 VHFQVGNFTRAHEAFVVEOIL-----EHPADWAIVGVLTCSDRSKKKAEFFKAQDCLYSL 84
QY 85 TETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMTITTEGGYNINETTGAF 144
Db 49 IEIAQKQOSRIEVTNAG--INSKEHQEIVIE-----AIQKTDIITTAIGPNI----- 94
QY 145 DLENAVKAADLKNPEKSTVFEGYVEAL-----RRWDAGGKAFVMSCDNLRHNGNVAR 199
Db 95 -----LPFIAELLAKGTEARRVAGNTQALDVMACENM-----IGG 129
QY 200 KAFGLYAKAR---DPELAKWIEENATFNGMVDRIPTVSAEIAKLNAASGLDDDLPLV 256
Db 130 SQFL-YQEVKKYLSPEGLTFADNYIGFPNAAVDRIVPTQSH-----DSLFVM 176
QY 257 AEDFHQWILEDOFADGRPPLEKAGVQMGVDVTWEYVKIRMLNAGHVMCLFCPGLVLGYEN 316
Db 177 VEPFENWVETKRLK-NPDLRLDEDVHYEEDLEPFIERKLFVSNSGHATSAYIGAHHYGAKT 235
QY 317 VDDATSELLGNLKNYLNDVIPTLKAP---SGMPTLEGYRDSVISRFSNKAMSDQTLRI 373
Db 236 ILEALQNPNIKSRIESVL-AEIRSLLIAKWNFDKKELENYHKVIIERFENPFIVDEVSrv 294
QY 374 ASDGSKVQVFTETVRRRAIEDKRLSRTAFGIAASYLEMLGRDEKGGTYESSEPTYGDA 433
Db 295 ARTPIRKLG--YNERFIRPIRELKLS-----LSYKNLLK-----TVG--- 330
QY 434 EWKLAKADDFESSLKLPAFGDWRDLTSELDOKVIVLRLKIIREKGVKAAI 483
Db 331 -----YAD-YRDVN-----DEESIRLGELLAKQSVKDVV 359

```

```

RESULT 5
US-09-815-242-5636
; Sequence 5636, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5636
; LENGTH: 368
; TYPE: PRN

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12393
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12393

Query Match
Best Local Similarity 21.4%; Score 181; DB 10; Length 368;
Matches 99; Conservative 64; Mismatches 168; Indels 132; Gaps 17;

QY 29 VHFVGNGFFRAHEAFVVEOILEHAPDWAIVGVLGTSRDKSKKAEFEKADCLYSLTETA 88
DB 4 VHFAGNIGRG-----FIGYIL-----ADNNVKVTFADVNEEIIINAL-AHDHQYDVI-LA 51
QY 89 PSKSTVRVMGALROYLLAPADPEAVLKHLYDPAIRIVSMITTEGGYININETTGAFDLEN 148
DB 52 DESKTTTRV-----NNVDA-----INSQPSSEALKQ 77
QY 149 AAVKADLNKPEKSTVFGYVVEALRRWDAGGKFTVMSCONLRHNGVARKAFGLGYAKA 208
DB 78 AILEADIIITAVGVNLPITIAKSFAPFLKEKTNHVNIVACENAIMATDTLKKAVL----- 132
QY 209 ROPELAKWTEENATPENGMDRITPTVSAETAKKLNAASGLDDDLPLVAEDFHQWVLEDO 268
DB 133 ---DITGPLGHNHIFANSADVRIPLQKNE-----NILDVMEPFYEWVEKD 177
QY 269 FADGRPLEKAGVQMGVDVTWMEYVKIRMLNAGHYMLCFPGILVGYENVDDAIEDSELLG 328
DB 178 AWYG-PELNH--IKYVDDLTPYIERKLLTVNTGHAYLAYAGKADKATVLDVAVKDSIEA 234
QY 329 NLKNYLKNDVIPTLAPSGMTLE-----GYRDSVISRFSNKMSDQTLRIASDGCS 379
DB 235 GURRLV-----AETSQYITNEFDTEAQAGYVEKIIDRFNNSYLSDEVTRVGRG--- 284
QY 380 KVOQFWTETVRRRAIEDKRDLRIAFGIASYLEMLRGRD-EKGGTYVESSEPTYGDAEWKLA 438
DB 285 -----TLRKIGPKDR-----IIKPLKLYNKDLERTGLLNTAA-----LLL 320
QY 439 KADDFESSLKLPAFGWRDLDTSELDQKVVILRKTIIEKGKYKA 481
DB 321 KYDD-----TADQETVEKNVYKEHGLKA 344

RESULT 7
US-09-932-145-7
; Sequence 7, Application US/09932145
; Patent No. US20020161191A1
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Mintier, Gabe
; APPLICANT: Kinney, Gene G
; APPLICANT: Ramanathan, Chandra S
; TITLE OF INVENTION: NOVEL IMIDAZOLINE RECEPTOR HOMOLOGS
; FILE REFERENCE: D0020 NP
; CURRENT APPLICATION NUMBER: US/09/932,145
; CURRENT FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1504
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: DATABASE ACCESSION NUMBER: Genbank,
; OTHER INFORMATION: accession number: NP_009115
US-09-932-145-7

Query Match
Best Local Similarity 21.2%; Score 125; DB 9; Length 1504;
Matches 102; Conservative 65; Mismatches 165; Indels 148; Gaps 25;
QY 23 GIKPGIV-HFVGNGNFFRAH--EAFYVEQITLEHAPDWAIVGVLGTSRDKSKKAEFEKADQ 79

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12393
; LENGTH: 368
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12393

Query Match
Best Local Similarity 21.4%; Score 181; DB 10; Length 368;
Matches 99; Conservative 64; Mismatches 168; Indels 132; Gaps 17;

QY 29 VHFVGNGFFRAHEAFVVEOILEHAPDWAIVGLTGSDRSKKAEFEKADCLYSLTETA 88
DB 4 VHFAGNIGRG-----FIGYIL-----ADNNVKVTFADVNEEINAL-AHDHQYDVI-LA 51
QY 89 PSKSTVRVMGALROYLLAPADPEAVLKHLYDPAIRIVSMITTEGGYNINETTGAFDLEN 148
DB 52 DESKTTTRV-----NNVDA-----INSQPSSEALKQ 77
QY 149 AAVKADLNKPEKSTVFGYVVEALRRWDAGGKFTVMSCONLRHNGVARKAFGLGYAKA 208
DB 78 AILEADIIITAVGVNLPITIAKSFAPFLKEKTNHVNIVACENAIMATDTLKKAVL----- 132
QY 209 ROPELAKWEEENATPENGMDRITPTVSAETAKKLNAASGLDDDLPLVAEDFHQWVLEDO 268
DB 133 ---DIITGPLGHNHIFANSADVRIPLQKNE-----NILDVMEFPFYEWNVEKD 177
QY 269 FADGRPLEKAGVQMGVDVTWMEYVKIRMLNAGHYMLCFPGILVGYENVDDAIEDSELLG 328
DB 178 AWYG-PELNH--IKYVDDLTPYIERKLLTVNTGHAYLAYAGKADKATVLDVAVKDSIEA 234
QY 329 NLKNYLKNDVIPLKAPSGMTLE-----GYRDSVISRFSNKMSDQTLRIASDGCS 379
DB 235 GURRLV-----AETSQYITNEFDTEAQAGYVEKIIDRFNNSYLSDEVTRVGRG--- 284
QY 380 KVOQFWTETVRRRAIEDKRLSRIAFGIASYLEMLRGRD-EKGGTYVESSEPTYGDAEWKLA 438
DB 285 -----TLRKIGPKDR-----IILKPLKLYNKDLERTGLLNTAA-----LLL 320
QY 439 KADDFESSLKLPAFGWRDLDTSELDQKVVILRKTIIEKGKYKA 481
DB 321 KYDD-----TADQETVEKNVYKEHGLKA 344

RESULT 7
US-09-932-145-7
; Sequence 7, Application US/09932145
; Patent No. US20020161191A1
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Mintier, Gabe
; APPLICANT: Kinney, Gene G
; APPLICANT: Ramanathan, Chandra S
; TITLE OF INVENTION: NOVEL IMIDAZOLINE RECEPTOR HOMOLOGS
; FILE REFERENCE: D0020 NP
; CURRENT APPLICATION NUMBER: US/09/932,145
; CURRENT FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1504
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: DATABASE ACCESSION NUMBER: Genbank,
; OTHER INFORMATION: accession number: NP_009115
US-09-932-145-7

Query Match
Best Local Similarity 21.2%; Score 125; DB 9; Length 1504;
Matches 102; Conservative 65; Mismatches 165; Indels 148; Gaps 25;
QY 23 GIKPGIV-HFVGNGNFFRAH--EAFYVEQITLEHAPDWAIVGLTGSDRSKKAEFEKADQ 79

```

```
Db 105 GTPRVLAHFLHFYEINGITAAALAELEFKEGQ--LLGAG-----EVFAIGP 151
Qy 80 C-LYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKHLVDPAIRIVSMFI--TEGGY- 135
Db 152 LQLTAVTEQLQQGAPT-----CASGDAKTDLGHILDFTCRLKYLKVSGETGPGF 200
Qy 136 --NINETTGAFDLENAAVKADLNKPEKPSVFGYVVEALRRRDAGGKAFVMSCDNLRH 193
Db 201 TSNIQEQLLPFDLS-----IFKSLHQ-----VEISHCDAKHI 232
Qy 194 NGNVARKAFGLYAKAR-----DPELAKWIEENATFPNGMVDRTITPTVSAEIA 240
Db 233 RGLVASPTLATLSVRFSATSMKEVLVPEASEFDEWEPEGTTL-EGPVTAVIPTWQALTT 291
Qy 241 KKL--NAASGLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVOMGVG-----TDW 290
Db 292 LDLSHNSISESVKLIP-----KIEFLD-----LSHNGLLVDNLQHLNVLVHLDL 339
Qy 291 EYVK-----IRMLN-AGHYMLCFFGI-----LVGYENVDDAIEDSELLGNLK 331
Db 340 SYNKLSSLEGLHTRKLNKIKTLNAGNLLESGLHLKLYSLVNLDRNRLEQMEEVRSIG 399
Qy 332 NYLNKDVIPITLKAPSGMTLEGYRDSVSRFSNKAAMS---DOTLRIADSGCKSVQVFWTET 388
Db 400 SLPCLHVSLLNNPLSI-IPDYRTKVLQAQGERASEVCLDDTDTVTTERE-LDTVEVL--KA 455
Qy 389 VRRAIEDKRDLSRIAFAGSIAYLEMLRGRDEKG-----TYESSEPTYGDAEWKL 437
Db 456 IOKAKEVSKLS-----NPEKKGEDSRLSAAPCTRPSSTPTVAPASASL 501

RESULT 8
US-09-824-574-3
; Sequence 3, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Mollanen, Anu-Maarit
; APPLICANT: Palvimio, Jorma J.
; APPLICANT: Jone, Olli A.
; TITLE OF INVENTION: ARIP4 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1703
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-824-574-3
```

```
Query Match 4.3%; Score 107; DB 9; Length 1703;
Best Local Similarity 21.5%; Pred. No. 2.5;
Matches 98; Conservative 75; Mismatches 168; Indels 114; Gaps 24;

Qy 45 VEQLEHAPDWAIVGVGLTSGDSRKKAEFEKADCLYSLTETAPSG-----KSTVRV 97
Db 280 IQKSINHPDFKRMLLSLEFARRRQPTDQ--NQSNLNGNGNTOOQGTNSHNTNTDNV 338
Qy 98 MGALRDYLL-----APADPEAV---LKLHVDPAIRIVS---MTITEGGYNINETT 141
Db 339 SGLTRNAPLDSKDENAFASVPAGSSVHNKNGTLDKNSQVSGTPTQTESKKEENET- 397
Qy 142 GAFDLENAVKA-----DLKNPEKPSVFGYVVEALRRRDAGGKAFVMSCD----- 189
Db 398 ----ISNVAKTAPNSKNKTHTEQNNPPKPQK--PVPNLVLDQOYKEGIVVDIDDPDMVD 451
Qy 190 -----NLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGMVDRTITPTVS 244
Db 452 SFTMPNISHS-NIDYQTLA-----NSDHAKFTTEPGVLPVG-IDTHTATDIYOTLIALN 504
```

```
Qy 245 AASGLDDDLPLV-----AEDFHQWLEDOFADGRPPLEKAGVOMGVGDVTDWEYVKIRMLN 299
Db 505 LDTTVNCLOKLNDCESTRENALYDYALQLPLQKA---VRGHVLOFENHONSLLT 561
Qy 300 AGHYMLCFFGILGYENVDDAIEDSELLGNLKNYLNKDVIPITLKAPSGMTLEGYRDSVIS 359
Db 562 NTH-----PNFLSKIRIN--VDALLTNQL--YKNHEL-----LKLERKKTEAVA 603
Qy 360 RFS--NKAMSDOTLRIASDGCKSVQVFWTETVRRRAIEDKRDLSRIAEG---IASYLEMLR 414
Db 604 RLKSMNKSALNOYNR-----RODKN-KRLKFGHRLIATHHTLER 642
Qy 415 GRDEKGYTESSEPTYGDAEWKLAKADDFESSLKL 449
Db 643 D-EQKRAEKRAKE-----RLQALKANDEEAYIKL 670

RESULT 9
US-09-801-368-340
; Sequence 340, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: NO. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 340
; LENGTH: 1703
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-340
```

```
Query Match 4.3%; Score 107; DB 10; Length 1703;
Best Local Similarity 21.5%; Pred. No. 2.5;
Matches 98; Conservative 75; Mismatches 168; Indels 114; Gaps 24;

Qy 45 VEQLEHAPDWAIVGVGLTSGDSRKKAEFEKADCLYSLTETAPSG-----KSTVRV 97
Db 280 IQKSINHPDFKRMLLSLEFARRRQPTDQ--NQSNLNGNGNTOOQGTNSHNTNTDNV 338
Qy 98 MGALRDYLL-----APADPEAV---LKLHVDPAIRIVS---MTITEGGYNINETT 141
Db 339 SGLTRNAPLDSKDENAFASVPAGSSVHNKNGTLDKNSQVSGTPTQTESKKEENET- 397
Qy 142 GAFDLENAVKA-----DLKNPEKPSVFGYVVEALRRRDAGGKAFVMSCD----- 189
Db 398 ----ISNVAKTAPNSKNKTHTEQNNPPKPQK--PVPNLVLDQOYKEGIVVDIDDPDMVD 451
Qy 190 -----NLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGMVDRTITPTVS 244
Db 452 SFTMPNISHS-NIDYQTLA-----NSDHAKFTTEPGVLPVG-IDTHTATDIYOTLIALN 504
Qy 245 AASGLDDDLPLV-----AEDFHQWLEDOFADGRPPLEKAGVOMGVGDVTDWEYVKIRMLN 299
```

Db 505 LDTTVNDCCLKLLNDECTESTRENALDYALQLLPLOKA---VRGHVLQFEWHQNSLLT 561
Qy 300 AGHVMLCPGILVGVYENVDATEDSELLGNLKNYLKNDVPTLKAPSGMTLEGYRDSVIS 359
Db 562 NTH-----PNFLSKTRNIN--VODALLNQL--YKNHEL-----LKLKKKTEAVA 603
Qy 360 RFS--NKAMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRLSRIAFG---IASYLEMLR 414
Db 604 RLKSNKSAINOYR-----RQDKKN-KRLKFGHRLIATHTNLER 642
Qy 415 GRDEKGGTYESEPTYGDAEWKLAADFFESSLKL 449
Db 643 D-EQRAEKAKE-----RLQALKANDEEAYIKL 670

RESULT 10

US-09-860-846-35
; Sequence 35, Application us/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860.846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-35

Query Match 4.2%; Score 105.5; DB 9; Length 1562;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;

Qy 87 TAPGSKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSTITEGGYNINET 140
Db 313 TAPHGPSOORVIRRALADARLAPGDVVEAHGTCTRLGDP-IEAQAIIATYVG-----QEK 367
Qy 141 TGAFDLENAAYKADLNKPEKSTVFGYV--VEALRR-----RWDAGGKA 182
Db 368 SSEQPLRLGALKSNIGHTQAAAGVAGVVKMVOAMRHGLLPKTLHVDEPSDQIDNSAGTVE 427
Qy 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWIEENATF--PNG-----MV 228
Db 428 LITEAVDPPEKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAVEPPAGGVVPPV 487
Qy 229 DRITP-TVSAETAKKLNAAAGLDDPLVAEDFHOWVLEDOFADGRPPLEKAGVOMVGDV 287
Db 488 SAKTPAALDAQIGQLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 538
Qy 288 TDWEYVKTRM-----LNAGHVMLCFPG-----ILVGVENVDDATEDSELLGNLKNY 333
Db 428 LITEAVDPPEKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAVEPPAGGVVPPV 487
Qy 229 DRITP-TVSAETAKKLNAAAGLDDPLVAEDFHOWVLEDOFADGRPPLEKAGVOMVGDV 287
Db 488 SAKTPAALDAQIGQLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 538
Qy 288 TDWEYVKTRM-----LNAGHVMLCFPG-----ILVGVENVDDATEDSELLGNLKNY 333
Db 539 REALDALRMEGLVGRGTSSDVGRVAFVFPQGTQWAGMGAEILLDSSPEFAASMAECETA 598
Qy 334 LNKDVIPITKA-----PSGMTLE-----GYRDSVISRFSNKAMSDQTL----- 371
Db 599 LSRVYDWSLEAVVREPQCAPTLDRVDVVPVTFVAVMSLAKVWQHGGITPQAVVCHSOG 658
Qy 372 -----RIASDGCCKVQVFWTETVRRRAIEDKRLSRIAFGIASYLEMLRGRDEKG--- 420
Db 659 IAAAYVAGALTLDAAARVVTLRKSIHAHLAKGKGMISLALDEAAVLRKLSDFDGLSVA 718
Qy 421 -----GTVESSEPTYGDAEWKLAADFFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIR 475
Db 719 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIIEKELA 770

Qy 476 E 476

Db 771 E 771

RESULT 11

US-09-988-384B-35
; Sequence 35, Application US/0988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988.384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-35

Query Match 4.2%; Score 105.5; DB 9; Length 1562;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;

Qy 87 TAPGSKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSTITEGGYNINET 140
Db 313 TAPHGPSOORVIRRALADARLAPGDVVEAHGTCTRLGDP-IEAQAIIATYVG-----QEK 367
Qy 141 TGAFDLENAAYKADLNKPEKSTVFGYV--VEALRR-----RWDAGGKA 182
Db 368 SSEQPLRLGALKSNIGHTQAAAGVAGVVKMVOAMRHGLLPKTLHVDEPSDQIDNSAGTVE 427
Qy 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWIEENATF--PNG-----MV 228
Db 428 LITEAVDPPEKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAVEPPAGGVVPPV 487
Qy 229 DRITP-TVSAETAKKLNAAAGLDDPLVAEDFHOWVLEDOFADGRPPLEKAGVOMVGDV 287
Db 488 SAKTPAALDAQIGQLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 538
Qy 288 TDWEYVKTRM-----LNAGHVMLCFPG-----ILVGVENVDDATEDSELLGNLKNY 333
Db 539 REALDALRMEGLVGRGTSSDVGRVAFVFPQGTQWAGMGAEILLDSSPEFAASMAECETA 598
Qy 334 LNKDVIPITKA-----PSGMTLE-----GYRDSVISRFSNKAMSDQTL----- 371
Db 599 LSRVYDWSLEAVVREPQCAPTLDRVDVVPVTFVAVMSLAKVWQHGGITPQAVVCHSOG 658
Qy 372 -----RIASDGCCKVQVFWTETVRRRAIEDKRLSRIAFGIASYLEMLRGRDEKG--- 420
Db 659 IAAAYVAGALTLDAAARVVTLRKSIHAHLAKGKGMISLALDEAAVLRKLSDFDGLSVA 718
Qy 421 -----GTVESSEPTYGDAEWKLAADFFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIR 475
Db 719 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIIEKELA 770

Qy 476 E 476

Db 771 E 771

RESULT 12

US-09-861-289-35

```
; Sequence 35, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; PRIOR FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-861-289-35

Query Match          4.2%; Score 105.5; DB 10; Length 1562;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;

Qy 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMITTEGGYNNINET 140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 TAPHGPSQORVIRRALADARLAPGDVVEAHGTCTRLGDP-IEAQAIIATYVG-----QEK 367
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 141 TCAFDLENAVKADLNKEKPTSTVFGYV--VEALRR-----RWDAGGKA 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 SSEQPLRLGALKSNIGHTQAAAGVAGVVKMVMQAMRHGLLPKTLHYDEPSDQIDWSAGTVE 427
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWIEENATF--PNG-----MV 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 428 LITEAVDPWPEKQDGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAEPAGGVVWPVPV 487
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 229 DRITP-TVSAEIAKLNAAAGLDDDLPLVAEDFHOWLEDQFADGRPPLEKAGVOMGDV 287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 488 SAKTPAALDAQIGLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 538
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 288 TDWEYVKIRM-----LNAGHVMLCFPG-----ILVGVENVDDAIEDSELLGNLKNY 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 REALDRLRMPEGLVRGTSSDVGRAVAFVPGCGTQWAGMGAEILLDSPEFAASMAECETA 598
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 334 LNKDVIPTLKA-----PSGMTLE-----GYRDSVISRFSNKAMSDOTL----- 371
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 599 LSRVYDWSLEAVVRQEPGAPTLDRVDVVPVPTFAVMVSLAKVMQHGGITPQAVVCHSOG 658
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 372 -----RIASDGCCKVOVFWETVTRRAIEDKRLDRIAFGIAASYLEMLRGRDEKG--- 420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 659 IAAAYVAGALTLDAAARVVTLRKSIAAHLAKGKGMISLALDEAAVLRKLSDFDGLSVAA 718
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 421 -----GTVESSEPTYGDAEWKLAKADDFESSLKLPFGDWRDLDTSELDQKVVILRKIIR 475
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 719 VNGPTATVVGDPQTQIEELARTCEADGVRARI-IP-----VDYASHRSRQVEIIIEKELA 770
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 476 E 476
|
Db 771 E 771

RESULT 13
US-09-860-846-6
; Sequence 6, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
```

```
; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-6

Query Match          4.2%; Score 105.5; DB 9; Length 11877;
Best Local Similarity 20.8%; Pred. No. 57;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;

Qy 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMITTEGGYNNINET 140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9001 TAPHGPSQORVIRRALADARLAPGDVVEAHGTCTRLGDP-IEAQAIIATYVG-----QEK 9055
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 141 TCAFDLENAVKADLNKEKPTSTVFGYV--VEALRR-----RWDAGGKA 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9056 SSEQPLRLGALKSNIGHTQAAAGVAGVVKMVMQAMRHGLLPKTLHYDEPSDQIDWSAGTVE 9115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWIEENATF--PNG-----MV 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9116 LITEAVDPWPEKQDGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAEPAGGVVWPVPV 9175
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 229 DRITP-TVSAEIAKLNAAAGLDDDLPLVAEDFHOWLEDQFADGRPPLEKAGVOMGDV 287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9176 SAKTPAALDAQIGLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 9226
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 288 TDWEYVKIRM-----LNAGHVMLCFPG-----ILVGVENVDDAIEDSELLGNLKNY 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9227 REALDRLRMPEGLVRGTSSDVGRAVAFVPGCGTQWAGMGAEILLDSPEFAASMAECETA 9286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 334 LNKDVIPTLKA-----PSGMTLE-----GYRDSVISRFSNKAMSDOTL----- 371
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9287 LSRVYDWSLEAVVRQEPGAPTLDRVDVVPVPTFAVMVSLAKVMQHGGITPQAVVCHSOG 9346
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 372 -----RIASDGCCKVOVFWETVTRRAIEDKRLDRIAFGIAASYLEMLRGRDEKG--- 420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9347 IAAAYVAGALTLDAAARVVTLRKSIAAHLAKGKGMISLALDEAAVLRKLSDFDGLSVAA 9406
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 421 -----GTVESSEPTYGDAEWKLAKADDFESSLKLPFGDWRDLDTSELDQKVVILRKIIR 475
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9407 VNGPTATVVGDPQTQIEELARTCEADGVRARI-IP-----VDYASHRSRQVEIIIEKELA 9458
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 476 E 476
|
Db 9459 E 9459

RESULT 14
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR FILING DATE: 09/105,537
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
```

QY 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMTITEGGYNINET 140

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:29:57 ; Search time 46 Seconds
(without alignments)
1013.591 Million cell updates/sec

Title: US-09-926-163B-2
Perfect score: 2516
Sequence: 1 MITRETLKSLPANVQAPPYD.....KVIVLRKIIREKGVKAAIPA 485

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922	36.6	491	2 D83353	mannitol dehydroge
2	789	31.4	502	2 S50519	hypothetical prote
3	779	31.0	487	2 A96022	probable fructuron
4	775.5	30.8	486	2 B91289	D-mannosate oxidor
5	775.5	30.8	486	2 E86130	mannonate oxidored
6	774.5	30.8	486	2 S56548	fructuronate reduc
7	771.5	30.7	486	2 A64909	probable fructuron
8	769.5	30.6	486	2 G90897	probable oxidoredu
9	768	30.5	478	2 T03548	mannitol 2-dehydro
10	763.5	30.3	486	2 H85719	probable oxidoredu
11	759.5	30.2	488	2 AD0679	fructuronate reduc
12	753.5	29.9	490	2 AF0156	probable D-mannona
13	746.5	29.7	488	2 B91011	probable oxidoredu
14	746.5	29.7	488	2 B95856	probable oxidoredu
15	739.5	29.4	488	2 C64986	probable fructuron
16	730	29.0	490	2 A80883	D-mannosate oxidor
17	730	29.0	494	2 AG3103	mannitol 2-dehydro
18	730	29.0	525	2 C98183	mannitol 2-dehydro
19	721	28.7	458	2 F87433	mannitol 2-dehydro
20	700.5	27.8	502	2 A53569	fructuronate reduc
21	650.5	25.9	490	2 A99293	mannitol dehydroge
22	650.5	25.9	490	2 A92990	mannitol dehydroge
23	569	22.6	463	2 AF0283	probable mannitol
24	371	14.7	539	2 H72422	hypothetical prote
25	365.5	14.5	544	2 H86828	fructuronate reduc
26	343.5	13.7	482	2 E96985	altronate oxidored
27	331	13.2	480	2 C89853	probable tagaturon
28	299.5	11.9	512	2 D83711	altronate oxidored
29	288	11.4	483	2 A80072	tagaturonate reduc

30	285	11.3	483	2 H90894	altronate oxidored
31	279	11.1	483	2 A85723	altronate oxidored
32	275	10.9	483	2 D64906	tagaturonate reduc
33	249.5	9.9	384	2 B96919	probable sigma fac
34	206	8.2	382	2 D91188	mannitol-1-phospha
35	206	8.2	382	2 D86035	mannitol-1-phospha
36	205	8.1	382	2 B65160	mannitol-1-phospha
37	202.5	8.0	387	2 AE0494	mannitol-1-phospha
38	193	7.7	374	2 C39435	mannitol-1-phospha
39	192	7.6	382	2 C44798	mannitol-1-phospha
40	191.5	7.6	385	2 H86627	hypothetical prote
41	190	7.6	366	2 I39888	mannitol-1-phospha
42	190	7.6	378	2 G97916	mannitol-1-phospha
43	188	7.5	378	2 H95045	mannitol-1-phospha
44	186	7.4	368	2 C90011	mannitol-1-phospha
45	185	7.4	382	2 AH0976	mannitol-1-phospha

ALIGNMENTS

RESULT 1
D83353
mannitol dehydrogenase PA2342 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83353
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83353
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-491 <STO>
A:Cross-references: GB:AE004660; GB:AE004091; NID:g9948372; PIDN:AAG05730.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: mtID; PA2342
C:Superfamily: conserved hypothetical protein YEL070w

Query Match	36.6%	Score 922;	DB 2;	Length 491;
Best Local Similarity	45.2%	Pred. No. 3.3e-57;		
Matches	199;	Conservative	66;	Mismatches 171;
			Indels	4;
			Gaps	2;
Qy	2	ITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILE--HAPDMAIVG	59	
Db	3	LNROHLPLLATAVARPSYDPAQLROGIVHGVGPHRAHAYTDALMNRGALDWAICG	62	
Qy	60	VLTGSDRSKKKAEFEKADCLYSLTETAPSGKSTVRVQMGALRDYLLAPADPEAVLKHLV	119	
Db	63	AGLRSDDRAMHDA--LAAQDYLYTLTGDPDPTVRVIGALSGMLLAEDGAELKLA	120	
Qy	120	DPAIRIVSMITTEGGYNINETTGAFDLENAAVKADLKNPKPSTVFYGVVEALRRWDAG	179	
Db	121	EPAIRIVSLITTEGGYCIDGSGEFLAELPLVRHDLANPRTPRGVGFLCEALRRRTDG	180	
Qy	180	GKFTVMSCDNLRHNGNVARKAFGLYAKARDELAKTTEENATFPNGMVDRTPTVSAEI	239	
Db	181	VPAFTVMSCDNLPHNGEVARKALLAFAERLDPGLARWIATHVSFPNAMVDRTPTSPAH	240	
Qy	240	AKKLNAAASGLDDPLVAEDFHQWLEDFADGRPPLEKAGVQWGVDTWDYEVKIRMLN	299	
Db	241	RRLQARHEDVEMPVVCEPFVQVQFVDFRSAGRPAWEKVGQVQFDDVTTPYEMKIGLLN	300	
Qy	300	AGHVMLCPFGILVGYENVDDAIEDSELNKNYLNKQDVIPTLKAPSGMTLEGYRDSVIS	359	
Db	301	GSHLALTYLGLRGYRFVHETLGDPLLRVYVAFMDRDVAPLLAPVPGIDILERYKDSLVE	360	
Qy	360	RFSNKAUSDQTLRIASDCGSKVQVFWTETVTRRAIEDKRLSRIAFGLSYLEMLRGRDEK	419	

Db 361 RFANRAIQDLERVCSDGSKFPKFIPTANRLTAAGRPLERVALVVAWALYLGVDHE 420
QY 420. GGTYESSEPTYGDAEWKLAK 439
Db 421 GERYPIPDRAAEQALVAE 440

RESULT 2
S50519
hypothetical protein YEL070w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N3810; hypothetical protein YNR073c
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 23-Mar-2001
C:Accession: S50519; S63406; S63411; S54055; S63890
R:Dietrich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of *S. cerevisiae* cosmids 9669, 8334, 8199, and lambda clone
A:Reference number: S50434
A:Accession: S50519
A:Molecule type: DNA
A:Residues: 1-502 <DIE>
A:Cross-references: EMBL:U18795; GSPDB:GN000005; MIPS:YEL070w; NID:g603241; PIDN:AAB65017
A:Genetics: C05
R:Andre, B.; Iraqui Houssaini, I.; Urrestarazu, L.A.; Vissers, S.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62920
A:Accession: S63406
A:Molecule type: DNA
A:Residues: 1-502 <AND>
A:Cross-references: EMBL:Z71688; GSPDB:GN000014; MIPS:YNR073c; NID:g1302609; PIDN:CAA9635
A:Experimental source: strain S288C
A:Genetics: C14
R:Levesque, H.; Lepingle, A.; Nicaud, J.M.; Gaillardin, C.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S63411
A:Accession: S63411
A:Molecule type: DNA
A:Residues: 1-438 <LEV>
A:Cross-references: EMBL:Z71688; GSPDB:GN000014; MIPS:YNR073c; NID:g1302609; PIDN:CAA9635
A:Experimental source: strain S288C
A:Genetics: C14
R:Levesque, H.; Nicaud, J.M.; Lepingle, A.; Gaillardin, C.
submitted to the EMBL Data Library, May 1995
A:Description: Sequence of a 9.2kb telomeric fragment from the right arm of *S. cerevisiae*
A:Reference number: S54055
A:Accession: S54055
A:Molecule type: DNA
A:Residues: 1-438 <LEV>
A:Cross-references: EMBL:X86790
A:Genetics: C14
R:Levesque, H.; Lepingle, A.; Nicaud, J.M.; Gaillardin, C.
Yeast 12, 289-295, 1996
A:Title: Sequencing of a 9.2 kb telomeric fragment from the right arm of *Saccharomyces cerevisiae*
A:Reference number: S63890; MUID:97060021; PMID:8904342
A:Accession: S63890
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-438 <LEV>
A:Cross-references: EMBL:X86790
A:Genetics: C14
A:Map position: 5L
C:Genetics: <C14>
A:Gene: MIPS:YNR073c
A:Map position: 14R
C:Superfamily: conserved hypothetical protein YEL070w

Query Match 31.4%; Score 789; DB 2; Length 502;
Best Local Similarity 35.7%; Pred. No. Re-48;
Matches 179; Conservative 90; Mismatches 195; Indels 38; Gaps 10;

QY 2 ITRETLSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILE--HAPDWAIVG 59

Db 12 LNAKTLKSFESTLPIPTYPREGVKGQIVHLGVGAFHRSHLAVFMHRLMQEHLHLDWSTCG 71
QY 60 VGLTGSORSKKAEFKAQDCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHV 119
Db 72 VGLMKADALMRDA--MKAQDCLYTLVERGIK-DTNAYIVGSITAYMAPDDPRAVIERMA 128
QY 120 DPAIRIVSMTITEGGYNINETTGAFDLENAAVKADLKNPEKPSVTFVGVVVEALRRRWDAG 179
Db 129 NPDTHIVSLTVTENGYYHSEATNSLMTDAPELINDLNHPEKPDLYGLYEAALLRYKRG 188
QY 180 GKAFVMSCDNLNRHNGVARKAFGLYAK-ARDPELAKWIEENATFPNGWVDRIPTPTVSAAE 238
Db 189 LTPFTINSCDNNPQNGVTVKTMVLVAFKKLDEKFAAWIEDKVTSPNSMVDVPTPRCTDK 248
QY 239 IAKKLNAASGLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVOMGVDTQWEXVKIRML 298
Db 249 ERKVVADWTGKIDQCPVVAEPFIOWLEDNFSQGRPVVQVQVVDVSELYMKLKL 308
QY 299 NAGHVMLCFPGILVGENVDDAIEDSELLGNLKNYLNKDVITPLKAPSGMTLEGYRDSVI 358
Db 309 NGHSAMGYLGLAGTYIHEVNDPTINKYIRVLMREEVPLLPKVPGVDFEEXTASVL 368
QY 359 SRFNSKAMSDDTLRIASDGSCKVQVFWTETVRRRAIEDKRD--LSRIARGIASYLEMLRGR 416
Db 369 ERFSNPAIQDTVARICLMSGKMPKYVLPISYEQLR-KPDGKYKLLAVCVAGWFRYLTGV 427
QY 417 DEKGTYESSEPTYGDAEWMKLAKADDFESSILKLPADFQWRD-----LDTSEL-DOK 466
Db 428 DMNGKPFIEIDP-----MAPTLKAAAVGCKDPHELLNIEVLFSPFIRDNK 473
QY 467 VIVLR-----KIIREKGVKAAI 483
Db 474 EFVAQLTHSLETVVDKGPAAI 495

RESULT 3
A96022
probable fructuronate reductase (EC 1.1.1.57) [imported] - *Sinorhizobium meliloti* (st
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001
C:Accession: A96022
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Her
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing e
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A96022
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-487 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49841.1; PID:g15141329; GSPDB:GN00167
A:Experimental source: strain 102L, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: uxuB; Smb20749
A:Genome: plasmid
C:Superfamily: conserved hypothetical protein YEL070w
C:Keywords: oxidoreductase

Query Match 31.0%; Score 779; DB 2; Length 487;
Best Local Similarity 38.9%; Pred. No. 3.9e-47;
Matches 166; Conservative 77; Mismatches 178; Indels 6; Gaps 3;

QY 2 ITRETLSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGVG 61

Db 3 LSRKTIIDRLPRTVKRPHYDLGTVTGIVHLGIGAFHRAHQAVYTDGLLSEDPGICGV 62
Qy 62 LTGDSRSKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALROYLLAPADPEAVLKLHLYDP 121
Db 63 L---RSPETRDALHPDGLYTLAVQDGE-SELSVGSVVELLCAPDPEAVLRRMADP 117
Qy 122 AIRIVSMITTEGGYINNETTGAFDLENAAVADLKNPKPSTVFGVYVEALRRRWADGCK 181
Db 118 GTRIVSLRTEKGYCHNPATGTLDEGHPDIVHDLANPARPSAIGFIVEAISRRYSAGIA 177
Qy 182 AFTVMSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAETAK 241
Db 178 PTLTSCDNLPGHVLKRIYVQFAEADPALAA-VWRNVASPTMVDRIYPAITDSRS 236
Qy 242 KLNAAAGLDDLLPLVAEDFHQWLEDDQADGRPPLEKAGQVMGVDVTWYVVKIRMLNAG 301
Db 237 AVASAMGLEDAWPIWTEPFROWIIEEDPLGRPAWEKAGALFVQDSAFEFMKLRLLNGS 296
Qy 302 HVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAQSGMTLEGYRDSVSRF 361
Db 297 HSTLAYGLYLAGAETVADAMALAGMEALVEGLMRHEVSTPLPELPGFDLPAYRAELQRF 356
Qy 362 SNKMSDQTLRIASDGCCKVQVFWETVTRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKGG 421
Db 357 RNPALRHRTWQIAMDGSOKLPORLLGSTRDLQAGAGYDRDLGVAAMWRYARGLDEAGR 416
Qy 422 TYESEP 428
Db 417 PIDVRDP 423
RESULT 4
B91289
D-mannanate oxidoreductase EC5282 [imported] - Escherichia coli (strain O157:H7, substr
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: B91289
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B91289
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA038705.1; PID:q13364760; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC5282
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.8%; Score 775.5; DB 2; Length 486;
Best Local Similarity 36.4%; Pred. No. 6.8e-47;
Matches 177; Conservative 87; Mismatches 205; Indels 17; Gaps 9;
Qy 9 SLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHA-PDWAIVGVGL-TGSD 66
Db 8 NLP--VARPSWDHSLRESRIVHLGCGAFHRAHQALYTHLLESTDSWGICEVNLMPGND 65
Qy 67 RSKKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALROYLLAPADP-EAVLKLHLYDPATRI 125
Db 66 R-VLIENLKKQQLLYTVAEKG-AESTELKIIIGSMKEALHPEIDGCEGILNAMARPOTAI 122
Qy 126 VSMITTEGGYINNETTGAFDLENAAVADLKNPKPSTVFGVYVEALRRRWADGCKAFTV 185
Db 123 VSLTVTEKGYCADAASGOLDLNNPLIKHDLNPTAPSAIGYIIVEALRLRREKGLKAPT 182
Qy 186 MSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAETAKLNA 245
Db 183 MSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAETAKLNA 242
Qy 246 ASGLDDDLPLVAEDFHQWLEDDQADGRPPLEKAGQVMGVDVTWYVVKIRMLNAGHYML 305
Db 243 QLVGYDPCAIACEPFROWIIEEDPLGRPAWEKAGALFVQDSAFEFMKLRLLNGS 302
Qy 306 CFPGLVGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAQSGMTLEGYRDSVSRFSNKA 365
Db 303 AYLGVLGGYETIADTWTNPYRKAFAFALMMQEOAPTLSMPEGTDLNAYATLLIERFSNPS 362
Qy 366 MSDQTLRIASDGCCKVQVFWETVTRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKGGTYES 425
Db 363 LRHRTWQIAMDGSOKLPORLLDVPVRLHONGGSRHRLALGVAGWMRYTQGVDEQGNADIV 422

Db 243 QLVGYDPCAIACEPFROWIIEEDNFVNGRPDWDKVGQAQVADVVPFEMMKLRMLNGSHSFL 302
Qy 306 CFPGLVGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAQSGMTLEGYRDSVSRFSNKA 365
Db 303 AYLGVLGGYETIADTWTNPYRKAFAFALMMQEOAPTLSMPEGTDLNAYATLLIERFSNPS 362
Qy 366 MSDQTLRIASDGCCKVQVFWETVTRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKGGTYES 425
Db 363 LRHRTWQIAMDGSOKLPORLLDVPVRLHONGGSRHRLALGVAGWMRYTQGVDEQGNADIV 422
Qy 426 SEPTYGDEWLAKADDPESLKLPAFGWGRDLDTSELDDQK-----VIVLRKIITREKV 479
Db 423 VDPML--AEFKINA-QYQAGRVKALLGLSGIFADDPQONADVFGAVTAAYQQLCERGA 479
Qy 480 KAAIPA 485
Db 480 RECVAA 485
RESULT 5
E86130
mannonate oxidoreductase [imported] - Escherichia coli (strain O157:H7, substrain EDL
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E86130
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86130
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <STO>
A:Cross-references: GB:AE005174; NID:q12519333; PIDN:AAG59505.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: uxuB
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.8%; Score 775.5; DB 2; Length 486;
Best Local Similarity 36.4%; Pred. No. 6.8e-47;
Matches 177; Conservative 87; Mismatches 205; Indels 17; Gaps 9;
Qy 9 SLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHA-PDWAIVGVGL-TGSD 66
Db 8 NLP--VARPSWDHSLRESRIVHLGCGAFHRAHQALYTHLLESTDSWGICEVNLMPGND 65
Qy 67 RSKKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALROYLLAPADP-EAVLKLHLYDPATRI 125
Db 66 R-VLIENLKKQQLLYTVAEKG-AESTELKIIIGSMKEALHPEIDGCEGILNAMARPOTAI 122
Qy 126 VSMITTEGGYINNETTGAFDLENAAVADLKNPKPSTVFGVYVEALRRRWADGCKAFTV 185
Db 123 VSLTVTEKGYCADAASGOLDLNNPLIKHDLNPTAPSAIGYIIVEALRLRREKGLKAPT 182
Qy 186 MSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAETAKLNA 245
Db 183 MSCDNLRHNGVARKAFGLYAKARDPELAKWIEENATFPNGMVDRIPTVSAETAKLNA 242
Qy 246 ASGLDDDLPLVAEDFHQWLEDDQADGRPPLEKAGQVMGVDVTWYVVKIRMLNAGHYML 305
Db 243 QLVGYDPCAIACEPFROWIIEEDNFVNGRPDWDKVGQAQVADVVPFEMMKLRMLNGSHSFL 302
Qy 306 CFPGLVGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAQSGMTLEGYRDSVSRFSNKA 365
Db 303 AYLGVLGGYETIADTWTNPYRKAFAFALMMQEOAPTLSMPEGTDLNAYATLLIERFSNPS 362
Qy 366 MSDQTLRIASDGCCKVQVFWETVTRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKGGTYES 425
Db 363 LRHRTWQIAMDGSOKLPORLLDVPVRLHONGGSRHRLALGVAGWMRYTQGVDEQGNADIV 422

Db 361 HRTWQIAMDGSKLPQRLMDSVRHLAHSKFDL--LALGVAGWMRYVGGVDEQGNPIEI 418
Qy 426 SEP-----TYGDAEWKLAKADDFESSKLPAFDGWRDLTSELDOKVIVLRKIIREK 477
Db 419 SDPLLVPVQKAVQSSAEGKAR-----VQSLLAIKAFGDDLPDNLSTFARVVTETVLSLLAH 474
Qy 478 GVKAAT 483
Db 475 GAKATV 480
RESULT 8
G90897
probable oxidoreductase ECS2151 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: G90897
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90897
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA835574.1; PID:g13361617; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: ECS2151
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.6%; Score 769.5; DB 2; Length 486;
Best Local Similarity 37.9%; Pred. No. 1.8e-46;
Matches 184; Conservative 78; Mismatches 201; Indels 23; Gaps 9;
Qy 10 LPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVEOI--LEHAPDMAIVGVLTGSDRS 68
Db 6 LSAKATLPPYDRNNLAPRIVHILGFAGPHRAHQGVYADILATEHFSDMGYEVEVNLIGGE-- 63
Qy 69 KKAAEFKAODCLYSLTETAPSGKSTVRMGALRDYLLAPADP--EAVLKLHVDPAIRTVS 127
Db 64 -QQTADLHQQNLTVAEMS--ADAWTARVGVVKKALHVQIDGLTVLAAMECFQOIALVS 121
Qy 128 MTITEGYNINETTGAFDLENAAVKADLNKPEKSTVFGYVVEALRRRDWAGGKRAFTVMS 187
Db 122 LTITEKGYFHSPTATQMLDHPMVAADVQNPQKATGTVIVEALARKAAGLPAFTVMS 181
Qy 188 CDNLHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWVDRTPTVTSAEIAKKLNAS 247
Db 182 CDNMPENGVNRDVTYSQAQVAVKLAQWIEDNVTFPSTWVDRIVPAVTEDTLAKIEOLT 241
Qy 248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEVVKIRMLNAGHVMCLF 307
Db 242 GVRDPAGVACPEPQWVEDNFVAGRPWEKAGALVSDVLPYEEMKURLMNGSHSELAY 301
Qy 308 PGILVGENVDDAIEDSELLGNLKNYLNKOVIPTLKPASGMTLEGYRDSVISRFSNKA 367
Db 302 LGYLAGYOHINCDMEDEHYHAAVALMLQEQAPTLLV--QGVVDLQDYANRLIARYSNPALR 360
Qy 368 DOTLRASDGSKVQVFTWTVR--RAIEDKRLSRIFAAGIASYLEMLRGRDEKGGYSES 425
Db 361 HRTWQIAMDGSKLPQRLMDSVRHLAHSKFDL--LALGVAGWMRYVGGVDEQGNPIEI 418
Qy 426 SEP-----TYGDAEWKLAKADDFESSKLPAFDGWRDLTSELDOKVIVLRKIIREK 477
Db 419 SDPLLVPVQKAVQSSAEGKAR-----VQSLLAIKAFGDDLPDNLSTFARVVTETVLSLLAH 474
Qy 478 GVKAAT 483
Db 475 GAKATV 480

RESULT 9
T03548
mannitol 2-dehydrogenase - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03548
R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromsome of Rhodobacter capsulatus SBI
A:Reference number: Z14955; MUID:97404404; PMID:9256491
A:Accession: T03548
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-478 <VLC>
A:Cross-references: EMBL:AF010496; NID:g9128256; PIDN:AACI6201.1; PID:g3128349
C:Genetics:
A:Map position: 1
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.5%; Score 768; DB 2; Length 478;
Best Local Similarity 36.4%; Pred. No. 2.2e-46;
Matches 176; Conservative 84; Mismatches 202; Indels 22; Gaps 9;
Qy 10 LPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVEOI--LEHAPDMAIVGVLTGSDR 67
Db 1 MPVPVAAAPRYDRAQLTPTGIVHILGNGFRAHQAVYLDLDFGLGEGHDWALLIGAGVRAPDA 60
Qy 68 SKKAAEFKAODCLYSLTETAPSGKSTVRMGALRDYLLAPADPEAVLKLHVDPAIRTVS 127
Db 61 AMR--EVLLAODGLSSVTELAP--GAHSARVIGAMIDFVAVQPGNAALTIATWARPEIRIVS 117
Qy 128 MTITEGYNINETTGAFDLENAAVKADLNKPEKSTVFGYVVEALRRRDWAGGKRAFTVMS 187
Db 118 LTVTEGYYIDPNTGTSPPDQIQAADARPDPATAGAIIAALKMRRRAAGVAPFTVMC 177
Qy 188 CDNLHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWVDRTPTVTSAEIAKKLNAS 247
Db 178 CDNYPHNGHVTDRVAVGLARLFDLAEADWIHASVAFNPMVDRTPT--ATGDRERAMAADL 236
Qy 248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEVVKIRMLNAGHVMCLF 307
Db 237 GLDDAAPVCTCTFFQWIEDKFPAGRPALAEKVGVTFTPHVDKFTMKIRILNGHAIIT 296
Qy 308 PGILVGENVDDAIEDSELLGNLKNYLNKOVIPTLKPASGMTLEGYRDSVISRFSNKA 367
Db 297 PGLADIATFVHEAMADPQIRAFNLRLTAELVLPVPPVPGDLDNDYKALIIIFERSNPEVA 356
Qy 368 DOTLRASDGSKVQVFTWTVRRAIEDKRLSRIFAAGIASYLEMLRGRDEKGGYSES 427
Db 357 DTVRLCLDGSNRQPKFTIPSIIRDRIATGGSFGLVLLSALWCRCYCFGTTESGAELAFND 416
Qy 428 PTYGDAEWKLAKADDFESSKLPAFDGWRDLT--SELDOKVIVLR-----KIIREKV 479
Db 417 PNW--ERLVAVA--QAQKDRPA--AWLEMEDIAEVRNPVAVIESFTTALKALWARGS 468
Qy 480 KAAI 483
Db 469 RAVI 472
RESULT 10
H85719
probable oxidoreductase ydfi [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85719
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamouis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85719
A:Status: preliminary

DB 6 / --TTLQELRXQDHLFSVLEKGAQGNQAI - VGSVCESVHARLDGIMQVLAKLVEPQVATV 122

Search completed: May 8, 2003, 16:34:29
Job time : 47 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:25:56 ; Search time 27 Seconds
(without alignments)

745.038 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516

Sequence: 1 MITRETLKSLPANVQAPPYD.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Length	ID	Description
1	789	31.4	502	1 YE10_YEAST	P39941 saccharomyc
2	774.5	30.8	486	1 UXUB_ECOLI	P3160 escherichia
3	771.5	30.7	486	1 YDFI_ECOLI	P77260 escherichia
4	739.5	29.4	488	1 YE10_ECOLI	P33029 escherichia
5	725.5	28.8	476	1 MTLK_RHOSH	P33216 rhodobacter
6	583	23.2	465	1 DALD_RALSO	P58708 ralstonia s
7	569	22.6	463	1 DALD_YERPE	P58709 versinia pe
8	534	21.2	455	1 DALD_KLEPN	O52720 klebsiella
9	354	14.1	228	1 YUUB_ERWHE	Q47826 erwinia her
10	285	11.3	483	1 UXAB_ECOLI	P24214 escherichia
11	258	10.3	50	1 POR_GLUOX	P80354 gluconobact
12	249.5	9.9	384	1 MTLD_CLOAB	O65992 clostridium
13	205	8.1	382	1 MTLD_ECOLI	P09424 escherichia
14	199	7.9	382	1 MTLD_BACST	Q45421 bacillus st
15	193	7.7	357	1 MTLD_ENTFA	P27543 enterococcu
16	192	7.6	382	1 MTLD_STRMU	Q02418 streptococc
17	190	7.6	366	1 MTLD_BACSU	P42957 bacillus su
18	184.5	7.3	374	1 MTLD_BACHD	Q9k681 bacillus ha
19	184	7.3	382	1 MTLD_KLEPN	Q9xbm6 klebsiella
20	138.5	5.5	385	1 MTLD_BUCAL	P7634 buchnera ap
21	123.5	4.9	691	1 EFG_BACSU	P80868 bacillus su
22	119	4.7	3685	1 DMD_HUMAN	P11532 homo sapien
23	118	4.7	3678	1 DMD_MOUSE	P11531 mus musculu
24	116.5	4.6	683	1 RPC_BPPHC	P08979 bacterioph
25	114	4.5	754	1 RPL1_INCUJ	P19703 influenza c
26	110	4.4	692	1 EFG_BACHD	Q92917 bacillus ha
27	109.5	4.4	3680	1 DMD_CANFA	O97592 canis fami
28	107	4.3	1703	1 SNF2_YEAST	P22082 saccharomyc
29	104.5	4.2	619	1 HSCA_HAEIN	P44669 haemophilus
30	103.5	4.1	775	1 YTX1_XENLA	P14380 xenopus lae
31	103	4.1	372	1 FTZL1_PYRAB	Q9v280 pyrococcus
32	103	4.1	675	1 VP55_YEAST	Q92331 saccharomyc
33	103	4.1	692	1 EFG_MYCPU	Q98qd8 mycoplasma

ALIGNMENTS

RESULT 1

ID	YE10_YEAST	-STANDARD:	PRT:	502 AA.
AC	P39941;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical 56.5 kDa protein in HXT8 5'region and in HXT17-COS10 intergenic region.			
DE	intergenic region.			
GN	YE1070W AND (YNR073C OR N3810).			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A. (YE1070W).			
RC	STRAIN=S288c / AB972;			
RA	Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,			
RA	Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,			
RA	Chung E., Duncan M., Guzman E., Hartzell G., Hunicke-Smith S.,			
RA	Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,			
RA	Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,			
RA	Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,			
RA	Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;			
RA	Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.			
RL	[2]			
RP	SEQUENCE FROM N.A. (YNR073C).			
RN	Andre B., Iraqi Houssaini I., Urrestarazu L.A., Vissers S.;			
RA	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.			
RL	[3]			
RP	SEQUENCE OF 1-439 FROM N.A. (YNR073C).			
RC	STRAIN=S288c;			
RX	MEDLINE=97060021; PubMed=8904342;			
RA	Levesque H., Nicaud J.-M., Lepingle A., Gaillardin C.;			
RT	"Sequencing of a 9.2 kb telomeric fragment from the right arm of			
RT	Saccharomyces cerevisiae chromosome XIV.";			
RL	Yeast 12:289-295(1996).			
CC	!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U18795; AAB65017.1; -			
DR	EMBL; Z71688; CAA96356.1; -			
DR	EMBL; Z71689; CAA96358.1; -			
DR	EMBL; X86790; CAA60486.1; -			
DR	SGD; S0000796; YEL070W			
DR	SGD; S0005356; YNR073C			
DR	InterPro; IPR0000669; Mannitol_dh.			
DR	Pfam; PF01232; Mannitol_dh; 1.			
DR	PRINTS; PR00084; MTLHDHGNASE.			

Q9cds1 lactococcus
O58698 pyrococcus
P12882 homo sapien
Q9v623 homo sapien
Q9ukx2 homo sapien
Q8umt1 methanosarc
P20735 sus scrofa
P43299 arabidopsis
Q9xdh6 pseudomonas
P36541 escherichia
Q10478 schizosacch
Q8zim7 versinia pe

34 103 4.1 877 1 DPO1_LACLA
35 103 4.1 967 1 SYL_PYRHO
36 103 4.1 1939 1 MYH1_HUMAN
37 103 4.1 1939 1 MYH4_HUMAN
38 103 4.1 1941 1 MYH2_HUMAN
39 102 4.1 568 1 DNLI1_METAC
40 101 4.0 568 1 GGT_PIG
41 101 4.0 716 1 PROL_ARATH
42 101 4.0 1173 1 DP3A_PSEPL
43 100.5 4.0 616 1 HSCA_ECOLI
44 100 4.0 475 1 YDF6_SCHPO
45 100 4.0 636 1 DNAK_YERPE

```
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Hypothetical protein; Oxidoreductase; NAD.
SQ SEQUENCE 502 AA; 56470 MW; A7E0CC01AADC1B2A CRC64;

Query Match 31.4%; Score 789; DB 1; Length 502;
Best Local Similarity 35.7%; Pred. No. 1.3e-46;
Matches 179; Conservative 90; Mismatches 195; Indels 38; Gaps 10;

QY 2 ITRTELKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFVVEOILE--HAPDWAIVG 59
DB 12 LNAKTLKSFESTLPIPTYPREGVKQGVHLGVGAFHRSHLAVFMHRLMQEHLKDWISICG 71
QY 60 VGLTSGDRSKKAEKFAQDCLYSTETAPSGKSTVVMGALRDYLLAPADPEAVLKHIV 119
DB 72 VGLMKADALMRDA--MKAQDCLYTLVERGIK-DTNAYIVGSITAYMYAPDDPRAVLEKWA 128
QY 120 DPAIRVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVGVYVEALRRWDAG 179
DB 129 NPDTHIVSLVTENGYYHSEATNSLMTDAPEIINDLNHPKPDTLGYLYEALLRYKRG 188
QY 180 GRAFTVMSCDNLRHNGNVARKAFGLYAK-ARPELAKWIEENATFPMGMVDRTTPTVSAE 238
DB 189 LPFTTMSCDNPNQGVYVTKMLVAFKLLKDEKFAWLEDKVTSPNSWDRVTPRCTDK 248
QY 239 IAKLNAAAGLDDDLPLVAEDFHQWLEDOFADGRPLEKAGVQMGVDYDWEYVKIRML 298
DB 249 ERKYADVATWGIKQDCPVAAEPFIQWLEDFNSDGRPPWELVGVQVVKDVSVELMKRL 308
QY 299 NAGHVMCLPFGILGVENVDADAIEDSELGNLKNYLNKDVIPTLKAPSGMTLEGYRDSVI 358
DB 309 NGHGSAMGYLGLAGTYIHEVNDPTINKYIRVLMREEVILPLPKVPGVDFEETASVL 368
QY 359 SRPSKAMSQDTRLIASDGCCKVQVFWTETVRRRAIEDKRD--LSRTAFGIASVLEMLRGR 416
DB 369 ERFNSPAIQDVTARICLMSGKMPKVLPSIYEQLR-KPDGKYKLLAVCVAGHFRYLTGV 427
QY 417 DEKGGTYESEPTGYDAEWKLAADDFESLKLPAFDGWRD-----LDTSEL-DOK 466
DB 428 DMNGKPFIEDP-----MAPTLKAAAVKGGKDPHELLNIEVLFSPEIRDKN 473
QY 467 VIVLR-----KIIREKGVKAAI 483
DB 474 EFVAQLTHSLETYDKGPAAI 495

RESULT 2
UXUB_ECOLI STANDARD; PRT; 486 AA.
AC P39160;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE D-mannonate oxidoreductase (EC 1.1.1.57) (Fructuronate reductase).
GN UXUB OR B4323
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RP [1]
RC SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RL region from 92.8 through 100 minutes.";
RN Nucleic Acids Res. 23:2105-2119(1995).
[2]
RC SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Mizobuchi K.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: D-mannonate + NAD(+) = D-fructuronate + NADH.
```



```
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
RN NCBI_TaxID=562;
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN NCBI_TaxID=562;
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kitagawa M., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Motomura K., Nakade S., Nakamura Y., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasundaram S.,
RA Yamamoto Y., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yanamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000251; AAC74615.1; -
DR EMBL; D90797; BAA15243.1; -
DR EMBL; D90798; BAA15247.1; -
DR EcoGene; EGI3821; ydgi.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHGNASE.
DR PROSITE; PS00974; MANNITOL_DHCENASE; 1.
DR KW Hypothetical protein; Oxidoreductase; NAD; Complete proteome.
DR NP_BIND 24 35 NAD (BY SIMILARITY).
DR SEQUENCE 486 AA; 53685 MW; 96F3BF07AE114B70 CRC64;
Query Match 30.7%; Score 771.5; DB 1; Length 486;
Best Local Similarity 37.7%; Pred. No. 1.9e-45;
Matches 183; Conservative 81; Mismatches 199; Indels 23; Gaps 9;
QY 10 LPANVQAPPYDIDGKPGIVHFGVGNFRAEAFVVEQI-LEHAPDWAIVGLVGSDBS 68
DB 6 LSAKATPLVLDNNAPRIVHFGAFHAFHOGVYADILATEHDFSDGWYVEVNLGGE-- 63
QY 69 KKAEEFKAQCLYSITETAPSGKSTVRVMGALRDYLLAPADP-EAVLKLHVDPAIRVS 127
DB 64 "QIADLOODNLYTVAES-ADWTVARVGVVKALHVOIDGLETVLAAMCEPOIAIVS 121
QY 128 MTITEGGYNINETTGAFLDENAANKADLNKPEKSTVFYGVVVEALRRWRDAGCKAFTVMS 187
DB 122 LTITEKGYFHSPTAQMLDHPMVAADYQNPQHPKTATGTVIVEALARRKAAGLPAFTVMS 181
QY 188 CDNLHNGNVARKAFGLYAKARDELPLAKWIENATFPNGWDRITPTVSAETAKLNAA 247
DB 182 CDMNPENGHVMDVVTYSQAQAVDKLAQWIEDNVTFPSTMDVRIYPVATEDFLAKIEQLT 241
QY 248 GLDDPLVAEDFHOWLEDOFADGRPPLEKAGVQMGVDGTDWEYVKIRMLNAGHVMCLCF 307
DB 242 GVRDPAGVACEPFRQWIEDNFVAGRPWEKAGAEVSDVLPYEEMKMLNGLNMGSHSFLAY 301
```

```
QY 308 PGILVGVENVDDAIEDSELLGNLKNLYNKDVITPTLKAPSGMTLEGYRDSVISRFSNKAMS 367
DB 302 LGYLAGYQHINDCMEDEHYRYAAYGLMLQEQAPTLKV-QGVLDQDYANRLIARYSNPALR 360
QY 368 DOTLRASDGCSKVQVFWTETVR--RAIEDKRDLRSIAFGIASYLEMLRGRDEKGGTYES 425
DB 361 HRTWQIAMDGSKLPQRLMDSVRWHLAHSKFDL--LAIGVAGMMRYGVGVDEQGNPIEI 418
QY 426 SEP-----TYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDQKVVILRKIREK 477
DB 419 SDPLLPVIOKAVOSSAEGKAR----VQSLLAIKAIFGDDLPDNLSTFARVTETYSLLAH 474
QY 478 GVKAAI 483
DB 475 GAKATV 480
RESULT 4
YEIQ_ECOLI STANDARD; PRT; 488 AA.
AC P33029; P94760;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical oxidoreductase yeiQ (EC 1.-.-.-).
OS YEIQ OR B2172.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasundaram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00007; AAA60520.1; -
DR EMBL; AE000306; AAC75233.1; -
DR EMBL; D90849; BAA15981.1; -
DR EcoGene; EGI2036; yeiQ.
```



```
RC STRAIN-CM11000;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Cholsne N., Claudel-Renard C., Cunnean S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- CATALYTIC ACTIVITY: D-arabinitol + NAD(+) = D-xylulose + NADH.
CC -!- PATHWAY: D-arabinitol catabolism; first step.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL646068; CAD15836.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRHGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; FALSE_NEG.
DR Oxidoreductase; NAD; Complete proteome.
DR KW
SQ SEQUENCE 465 AA; 50630 MW; 47095425D4836342 CRC64;

Query Match 23.28; Score 583; DB 1; Length 465;
Best Local Similarity 35.98; Pred. No. 1.3e-32;
Matches 142; Conservative 63; Mismatches 177; Indels 14; Gaps 8;

Qy 26 PGIVHGVGNFFRAHEAFYVBOILEHAPD---WAIIVGVGLTSGDRSKKKAEEFKAQDCLY 82
Db 10 PVLHIGAGSFHRAHQAWYLRHVNAVPPGERWILT-VGNIRDDHWTALA-ALAAQQGAY 67

Qy 83 SLTEPAPSGKSTVRVMGALRDYLLAPADPEAVLKHVDPAIRIVSMITTEGGYNINETTG 142
Db 68 TLETVTPOGERAYETIRSIARVLPWSAALINTGADPACRIVSFTVTTEGGYILDE-HD 126

Qy 143 AFDLENAVAKADLNKPEKPSIVFGVVEALRRWDAGKATVMSCDNLRHNG---NVAR 199
Db 127 RDVTHPDLAALRGAR--STLYGALAALLAERRRGAGPLTLOSCDNLNRNGARFRAGM 184

Qy 200 RAFLGYAKARDPELAKWIEENATFPNGMVDRTPTVSAEIAKKLNAASGLDDDLPLVAED 259
Db 185 RAFL--ALRGDAALLAWFEDANVSCPSAMVDRTPTDDVTRVHAATGVDDRCFVGES 242

Qy 260 FQWVLEDOFADGRPLEKAGVQVGDVTDWEYVKIRMLNAGHVMCLPFGILVGVYVDD 319
Db 243 FQWVIEDNFIAGRPAWEIAGAEIVADVHPVEAKIRILNATHSCIAWAGTLAGLYIHE 302

Qy 320 AIEDSELLGNLKNYLNKQVITPLKAPSCWTLGEGYRDSVISRFSNKAMSQDQILRTASDCS 379
Db 303 GMRDAATIRFAYDVYDDVIVCL-TPSPDLLERYVDVYLERFNGPNVYLDITNORVAADGFS 361

Qy 380 KVQFWETVTRRAIEDKRDLSRIAPGSIAYLEMLRG 415
Db 362 KIPGIATPLAECAFARGADPVATVLPALFLGFLG 397

RESULT 7
DALD_YERPE STANDARD; PRT; 463 AA.
AC P58709;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE D-arabinitol 4-dehydrogenase (EC 1.1.1.11).
GN DALD OR YPO2325.
OS Yersinia pestis.
```

```
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OC NCBI_TaxID=632;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / Biovar Orientalis;
MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jageis K., Karylshv A.V.,
RA Leather J., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -!- CATALYTIC ACTIVITY: D-arabinitol + NAD(+) = D-xylulose + NADH.
CC -!- PATHWAY: D-arabinitol catabolism; first step.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ414152; CAC91130.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PROSITE; PS00974; MANNITOL_DHGENASE; FALSE_NEG.
DR Oxidoreductase; NAD; Complete proteome.
DR KW
SQ SEQUENCE 463 AA; 51984 MW; D82E988685B5EDD CRC64;

Query Match 22.68; Score 569; DB 1; Length 463;
Best Local Similarity 36.08; Pred. No. 1.2e-31;
Matches 133; Conservative 71; Mismatches 145; Indels 20; Gaps 10;

Qy 29 VHFVGNGFFRAHEAFYVBOILEHAPD-WAIIVGVGLTSGDRSKKKAEEFKAQDCLYSLTET 87
Db 13 LHIGAGSFHRAHQAWYLRHLASGDDRTIALANI--RDDAIPLDLTLRAHQGEVLETV 70

Qy 88 APSKSTVRVMGALRDYLLAPADPEAVLKHV---DPAIRIVSMITTEGGYNINETTG 143
Db 71 TPSEGRQYKETSILRN--IIPWD--KALDHLVAQSGKPETRVISFTVTEGGYILD---NQ 123

Qy 144 FDL--ENAAVAKADLNKPEKPSIVFGVVEALRRWDAGKATVMSCDNLRHNGNVARKA 201
Db 124 FNLQODNSDIOADLRGDCR--TIYGASIRLQORQODKSGPVTLNCDNLRHNGERFRHG 181

Qy 202 FLGYAKARDPE-LAKWIEENATFPNGMVDRTPTVSAEIAKKLNAASGLDDDLPLVAEDF 260
Db 182 LLEFLALRGQSLNWNVTISQTRSPNTWVDRTTPRSADIARVLEKTGIVDNAPVWGSEF 241

Qy 261 HOWVLEDOFADGRPLEKAGVQVGDVTDWEYVKIRMLNAGHVMCLPFGILVGVYVDDA 320
Db 242 IQWVIEDNFIAGRPALENVGAEMVTSVLPYBEAKIRILNASHSCIAWAGSLIGQCYIHES 301

Qy 321 IEDSELLGNLKNYLNKQVITPLKAPSCWTLGEGYRDSVISRFSNKAMSQDQILRTASDCSK 380
Db 302 TOTDAIRQOMAYDYVTDVIPS-L-TPSPDLALAHYRDVLERFSNPYIRDTNORVAADGFSK 360

Qy 381 VOVEWETV 389
Db 361 IPGFITPTL 369

RESULT 8
DALD_KLEPN STANDARD; PRT; 455 AA.
AC O52720;
DT 15-JUN-2002 (Rel. 41, Created)
```

```
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE D-arabinitol 4-dehydrogenase (EC 1.1.1.11).
GN DALD.
OS Klebsiella pneumoniae.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCBI_TaxID=573;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1033-5p14 / KAY2026;
RX MEDLINE=98304087; PubMed=9639934;
RA Heuel H., Shakeri-Garakani A., Turgut S., Lengeler J.W.;
RT "Genes for D-arabinitol and ribitol catabolism from Klebsiella
   pneumoniae."
RL Microbiology 144:1631-1639(1998).
CC -!- CATALYTIC ACTIVITY: D-arabinitol + NAD(+) -> D-xylulose + NADH.
CC -!- PATHWAY: D-arabinitol catabolism; first step.
CC -!- SUBUNIT: Monomer.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF045245; AAC26498.1; .
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDDHGRNASE.
DR PROSITE; PS00974; MANNITOL_DHEGENASE; FALSE_NEG.
KW Oxidoreductase; NAD.
SQ SEQUENCE 455 AA; 51022 MW; 2D446E5232778DD5 CRC64;

Query Match      21.2%; Score 534; DB 1; Length 455;
Best Local Similarity 31.1%; Pred. No. 2.8e-29;
Matches 114; Conservative 85; Mismatches 162; Indels 6; Gaps 5;

QY 29 VIFGCVGNFRAHEAFYVEQILEHAPDPAVLEHVPALRIYSMTITTEGYNNINETTGAFDLEN 148
DB 8 LHIGLGSFHRAHQWYHLRLIASGDNWRIRIAAGNIRND-AEQVQALAAQGGRYVLETVS 66
QY 89 PSKSTVRVMGALRDYLLAPADPEAVLKLHVPALRIYSMTITTEGYNNINETTGAFDLEN 148
DB 67 PGEREYEITSIQKLLPWQAGLQPLINEGANPQTKVIAFTVTEGGYILN-TRHRETSN 125
QY 149 AAVKADLNKPKSTPVFGYVVEALRRWRDAGGKFTVMSCDNLRHNGVNVARKAFLGYAKA 208
DB 126 PDLQADLQGECK--TIYGTFLARILEKRMADNAGPLTLNCDNVRHNGERFDHGMVEFLQL 183
QY 209 RDEPE-LAKWIENATFPNGWDRIPPTVSAEATKLNAAAGSLDDDLPLVAEDFHQWVLED 267
DB 184 TGQAVVIDMMAATCPTNTMVDTRIPRPAADLPARIKAQAGIDDKAPVMGETTFIOMVVEN 243
QY 268 QFADGRPPLEKAGVQMGVDTWEYVKIRMLNAGHVMLCFPGILVGYENVDDAIEDSELL 327
DB 244 NFRDVRPNLEAVGVEMVESVIPEEKIRLNASHSCIAWAGTLICQQYIHSTLTDVIY 303
QY 328 GNLKYNLKNIDVPTLKAPSGMTLEGYRDSVISRFSNKAMSDQTLRIASDCGSKVQVFWTE 387
DB 304 AIADRVYTEDVIPCL-GDNGIDLPYRDVVLKRFNTNPIYQDTRNQVAADGFSKIPAMIAPI 362
QY 388 TVRRATE 394
DB 363 TLQECYQ 369

RESULT 9
YTUB_ERWHE STANDARD; PRT; 228 AA.
ID YTUB_ERWHE
```

```
Q47826;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical protein in tubB 3'region (Fragment).
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 21434;
RA Poor F.;
RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U25347; AAA66392.1; .
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PROSITE; PS00974; MANNITOL_DHEGENASE; PARTIAL.
KW Hypothetical protein; Oxidoreductase; NAD.
FT NP_BIND 48 59 NAD (BY SIMILARITY).
FT NON_TER 228 228
SQ SEQUENCE 228 AA; 24588 MW; 0DC5756A85FDF410 CRC64;

Query Match      14.1%; Score 354; DB 1; Length 228;
Best Local Similarity 41.5%; Pred. No. 2.1e-17;
Matches 85; Conservative 28; Mismatches 86; Indels 6; Gaps 4;

QY 8 KSLPANVOAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQI-LEHAPDPAVLEHVPALRI 66
DB 28 KLLSANATVPQWDRHNLVPRIVHLGFGAFHRAHQAVYADVLAAEHSSDWGYIEVNLIGGE 87
QY 67 RSKKKAEEFKAQDCLYSLTETAPSKSTVRVMGALRDYLLAPAD-PEAVLKLHVPALRI 125
DB 88 ---QQIADLKCODEFLTYVAEMSASWSG-RVVGVRQALHAQVDTGEAVLEALCQPVAI 143
QY 126 VSMITTEGYNNINETTGAFDLENAAVKAADLNKPKSTPVFGYVVEALRRWRDAGGKFTV 185
DB 144 VSLTITEKGYCHAPATGELMRDHPPLIVADLAQPHQPHSAPGLIIVELARRRAAGLAASFV 203
QY 186 MSCDNLNRHNGVNVARKAFLGYAKARD 210
DB 204 MSCDNIPENGRVAGVVCAYARQCD 228

RESULT 10
UXAB_ECOLI STANDARD; PRT; 483 AA.
ID UXAB_ECOLI
AC U24214; P78064;
DT 01-MAR-1992 (Rel. 21, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Altronate oxidoreductase (EC 1.1.1.36) (Tagaturonate reductase)
DE (Tagaturonate dehydrogenase).
GN UXAB BI521 OR Z2184 OR ECS2128.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA M12obuch1 K.;
```

Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattnner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12";
RL Science 277:1453-1474(1997).
[3]
SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
Kasai H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M.,
Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T., Saito N.,
Sampel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map";
RL DNA Res. 3:363-377(1996).
[4]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouls K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
RL Nature 409:529-533(2001).
[5]
SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
[6]
SEQUENCE OF 1-45 FROM N.A., AND SEQUENCE OF 1-4.
RC STRAIN=K12;
RX MEDLINE=86280350; PubMed=3525742;
RA Blanco C., Mata-Gilsinger M.;
RT "A DNA sequence containing the control sites for the *uxaB* gene of
Escherichia coli";
RL J. Gen. Microbiol. 132:697-705(1986).
CC -!- CATALYTIC ACTIVITY: D-altronate + NAD(+) = D-tagaturonate + NADH.
CC -!- PATHWAY: Glucuronate pathway.
CC -!- INDUCTION: BY GALACTURONATE, TAGATURONATE AND FRUCTURONATE. ITS
CC EXPRESSION IS SUBJECTED TO CATABOLITE REPRESSION BY GLUCOSE.
CC -!- SIMILARITY: SOME, TO THE MANNITOL DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D13327; BAA02586.1; -
CC EMBL; AE000249; AAC74594.1; -
CC EMBL; D90795; BAA15204.1; -
CC EMBL; AE005352; AAG56245.1; -

DR EMBL; AP002557; BAB35551.1; -
DR EMBL; M15737; AAA24757.1; -
DR PIR; A46549; A46549.
DR ECoGene; EG11065; uxAB.
DR InterPro; IP000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
KW Oxidoreductase; NAD; Complete proteome.
FT CONFLICT 332 332. P -> L (IN REF. 2).
SQ SEQUENCE 483 AA; 54808 MW; AFA4677CF35BC2E0 CRC64;

Query Match 11.38; Score 285; DB 1; Length 483;
Best Local Similarity 23.68; Pred. No. 3.le-12;
Matches 120; Conservative 89; Mismatches 200; Indels 100; Gaps 21;

Qy 28 IVHFGVGNFFRAHFAFYVEQILEHAP-DWAIVGVGLTGSRSKKKAE-----EFAKQDCL 81
Db 18 IIOFGEGNLFRAFVQMDIDLLNEHTDLNSGVVV-----RPIETSFPPSLSTQDGL 68
Qy 82 YS-----LTETAPSKSTVRVMGALROYLLAPADPEAVLKHLDPAIRIVSWTITEGGYN 136
Db 69 YTTIRGLNEKG-EAVSDARLIRSVMREISVSEYDEFLKLAHNPDMRFVFSNTTEAGIS 127
Qy 137 INEYTGAFDLENAAVKADLKNPEKPSYVGVVEALRRRWAGGKAFVMSCDNLRHNGN 196
Db 128 YH-AGDKFD-DAPAVSYPAKUTRLLEFERSHFNAGALDKG-----IIIPCGLIDYNGD 178
Qy 197 VARRAFLGYAK--ARDPELAKWIEENATFPNGMVDRIPTVSAETAKKLNAAAGLDDDL 254
Db 179 ALRELVLRYAQEWALPEAFIOWLOANSFCSTLVLDRIYVTPRDEVAKLEELGYHDGFL 238
Qy 255 LVAEDFHQWLE-----DOFADGRPLEKAGVOMYGVDTVDWEYVYKIMLNAGHV 303
Db 239 DTAEHFYLFTVQGPKSALTELRLDKY-----PL-----NVLLVDDIKPKYKRVAILNGAHT 290
Qy 304 MLCPPGILVGVENVDDAIEDSELNKNLYLNKQVITPKAPSGMTLCGYRSDSVLSRSEN 363
Db 291 ALVPVAFQAGLDTGEAMNDAEICAFVEKATYEEIIPVLDLPRD-ELESFASAVTGRFN 349
Qy 364 KAMSDQTLRIASDGCCKVQVFWETVTRRAIEDKRD-----LSRIAFGIASYLEMLRGREK 419
Db 350 PYIKHQLLSIALNGWK---FTRILPOLLGAKANGTLPARLTFALAALLATFYRG-ERN 405
Qy 420 GGTVESSEPTVGDREW-----KLAKADDFESSL-KLPAPFG 454
Db 406 GETY-----PVQDDAHLERYQQLWSQHRDRVIGTQELVAIVLAEKDHWEDLTQVP---G 458
Qy 455 WRDLDTSELDOKVIVLRKIIIEKGVKAAI 483
Db 459 LVEQVANDLD-----AILEKGMREAV 479

RESULT 11
POR_GLUOX
ID POR_GLUOX STANDARD; PRT: 50 AA.
AC P80354;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polyol:NADP oxidoreductase (EC 1.1.1.-) (Fragment).
GN POR.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE.
RC STRAIN=DSM 350;
RA Klaseen R.;
RL Thesis (1994), Heinrich-Heine University / Dueseldorf, Germany.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
DR InterPro: IP000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
KW Oxidoreductase; NADP.


```

Query Match          7.7%  Score 193;  DB 1;  Length 357;
Best Local Similarity 21.7%  Pred. No. 3.9e-06;
Matches 89; Conservative 65; Mismatches 144; Indels 112; Gaps 17;

QY      29  VHFVGNGFFRAHEAFYEQILEHAHPDWAIVGVGLTSDRSKKKAEERKAODCLYSLTETA 88
      ||||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db       4  VHEGAGNIGRG-----AKTGFI LPPFVDVNGNHHQALKER----- 43

QY      89  PSGKSTVRVMGALRDYLLAPADPEAVLKHLVDPAIRIVSMTITEGGYNINETTGAFDLEN 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      44  -----KSYTIETELADAS-----HQGINVENVT-----GLNHNMTPEKVVE 77

QY     149  AAVKADKLKNPEKPTVFYGVYVEALRRRRWDAGGKA-----FTVMSCDNLNRHNGNVARKAF 202
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:28:22 ; Search time 38 Seconds
(without alignments)
2629.812 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516

Sequence: 1 MITRETLKSLPANVOAPPDY.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_21.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phase.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertibrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2516	100.0	485	2 Q9KWR5	Q9kwr5 gluconobact
2	2143	85.2	485	2 Q8VUU5	Q8vuu5 gluconobact
3	922	36.6	491	16 Q91ID6	Q91id6 pseudomonas
4	888	35.3	493	2 O8B355	O8b355 pseudomonas
5	797	31.7	492	16 Q98D06	Q98d06 rhizobium l
6	779	31.0	487	16 Q92TQ9	Q92tq9 rhizobium m
7	775.5	30.8	486	16 Q8X516	Q8x516 escherichia
8	769.5	30.6	486	16 Q8X506	Q8x506 escherichia
9	768	30.5	478	2 O68111	O68111 rhodobacter
10	760	30.2	494	16 Q92N07	Q92n07 rhizobium m
11	759.5	30.2	488	16 Q8Z626	Q8z626 salmonella
12	755.5	29.9	480	16 Q8ZP18	Q8zpi8 salmonella
13	753.5	29.9	490	16 Q8ZGK9	Q8zgk9 yersinia pe
14	746.5	29.7	488	16 Q8XE83	Q8xe83 escherichia
15	730	29.0	490	16 Q8XFB5	Q8xfb5 salmonella
16	730	29.0	525	16 Q8U7J8	Q8u7j8 agrobacteri

17	721	28.7	458	16 Q9A877	Q9a877 caulobacter
18	700.5	27.8	502	16 Q8YCO0	Q8ycq0 brucella me
19	678.5	27.0	499	16 Q8EWM0	Q8ewo0 rhizobium l
20	650.5	25.9	490	16 Q8UA45	Q8ua45 agrobacteri
21	587.5	23.4	490	16 Q8ZM58	Q8zm58 salmonella
22	556.5	22.1	497	16 Q98JL8	Q98jl8 rhizobium l
23	518.5	20.6	439	2 Q9XBD5	Q9xbd5 amycolatops
24	405	16.1	539	16 Q8R8Q5	Q8rg5 thermoanaer
25	392.5	15.6	285	2 Q93KC4	Q93kc4 erwinia chr
26	371	14.7	539	16 Q9WXS3	Q9wxs3 thermotoga
27	365.5	14.5	544	16 Q9CF49	Q9cf49 lactococcus
28	343.5	13.7	482	16 Q97L67	Q97l67 clostridium
29	331	13.2	480	16 Q34354	Q34354 bacillus su
30	299.5	11.9	512	16 Q9KFI7	Q9kfi7 bacillus ha
31	288	11.4	483	16 Q8ZIC5	Q8zic5 yersinia pe
32	238	9.5	386	16 Q8RCS0	Q8rcs0 thermoanaer
33	214	8.5	366	2 Q9KH89	Q9kh89 escherichia
34	213	8.5	366	2 Q9KH91	Q9kh91 escherichia
35	209	8.3	366	2 Q9KH92	Q9kh92 escherichia
36	208	8.3	366	2 Q9K2M3	Q9k2m3 escherichia
37	207	8.2	366	2 Q9K2H8	Q9k2h8 escherichia
38	206	8.2	366	2 Q9KH90	Q9kh90 escherichia
39	206	8.2	382	16 Q8XDG9	Q8xdg9 escherichia
40	205	8.1	366	2 Q9K2Y2	Q9k2y2 escherichia
41	205	8.1	366	2 Q9K202	Q9k2u2 escherichia
42	202.5	8.0	387	16 Q8Z9X0	Q8z9x0 yersinia pe
43	202	8.0	366	2 Q9K2V6	Q9k2v6 escherichia
44	202	8.0	366	2 Q9KH93	Q9kh93 escherichia
45	201	8.0	366	2 Q9KH88	Q9kh88 escherichia

ALIGNMENTS

RESULT 1

Q9KWR5 ID Q9KWR5 PRELIMINARY; PRT; 485 AA.
AC Q9KWR5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 18, Last annotation update)
DE Sorbitol dehydrogenase.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G624;
RA Shibata T., Ichikawa C., Matsuura M., Takata Y., Noguchi Y., Saito Y., Yamashita M.
RT "Cloning of a gene for D-Sorbitol dehydrogenase from Gluconobacter oxydans G624 and expression of the gene in Pseudomonas putida IF03738".
RL J. Biosci. Bioeng. 89:463-468(2000).
DR EMBL; AB028937; BAA99414.1; .
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDHGNASE.
SQ SEQUENCE 485 AA; 53642 MW; A7C363D728EEA7DC CRC64;

Query Match 100.0%; Score 2516; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.5e-173;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MITRETLKSLPANVOAPPDYIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
Db 1 MITRETLKSLPANVOAPPDYIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
Qy 61 GLTGSDRSKKAEERKADCLYSLTETAPSGKSTVRVMGALRDYLLAPADPAVLKHLVD 120
Db 61 GLTGSDRSKKAEERKADCLYSLTETAPSGKSTVRVMGALRDYLLAPADPAVLKHLVD 120

```
QY 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKPTSTVFYVVEALRRRWDAGG 180
|||||
Db 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKPTSTVFYVVEALRRRWDAGG 180

QY 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSAETA 240
|||||
Db 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSAETA 240

QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWYVKIRMLNA 300
|||||
Db 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWYVKIRMLNA 300

QY 301 GHVLMCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360
|||||
Db 301 GHVLMCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360

QY 361 FSNKAMSDOTLRIASDGCSCVQVFTETVRRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
|||||
Db 361 FSNKAMSDOTLRIASDGCSCVQVFTETVRRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420

QY 421 GYSESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480
|||||
Db 421 GYSESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480

QY 481 AAIPA 485
|||||
Db 481 AAIPA 485
```

RESULT 2

```
Q8VUU5 ID Q8VUU5 PRELIMINARY; PRT; 485 AA.
AC Q8VUU5;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE L-sorbose reductase.
GN SR.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO3291;
RX MEDLINE=21650687; PubMed=11790761;
RA Shiojoh M., Tazoe M., Hoshino T.;
RT "NADPH-Dependent L-Sorbose Reductase Is Responsible for L-Sorbose
RT Assimilation in Gluconobacter suboxydans IFO 3291.";
RL J. Bacteriol. 184:861-863(2002).
DR EMBL; AB063188; BAB83933.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
SQ SEQUENCE 485 AA; 53541 MW; 883EC4956B3C6CF0 CRC64;
```

Query Match 85.2%; Score 2143; DB 2; Length 485;
Best Local Similarity 84.5%; Pred. No. 1.3e-146;
Matches 410; Conservative 30; Mismatches 45; Indels 0; Gaps 0;

```
QY 1 MITRETLKSLPANVQAPPYDIDGIRKPGIVHFGVGNFFRAHEAFYVEQILEHAPDAIVGV 60
|||||
Db 1 MITRETLKSLPANVQAPPYDIDGIRKPGIVHFGVGNFFRAHEAFYVEQILEHAPDAIVGV 60

QY 61 GLUTGDSRSKKKAEKFAKQDCLSLTETAPSGKSTVRVWGALRDYLLAPADPEAVLKLVD 120
|||||
Db 61 GLUTGDSRSKKKAEKFAKQDCLSLTETAPSGKSTVRVWGALRDYLLAPADPEAVLKLVD 120

QY 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKPTSTVFYVVEALRRRWDAGG 180
|||||
Db 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKPTSTVFYVVEALRRRWDAGG 180

QY 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSAETA 240
|||||
```

```
Db 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSAETA 240
|||||
QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWYVKIRMLNA 300
|||||
Db 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWYVKIRMLNA 300

QY 301 GHVLMCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360
|||||
Db 301 GHVLMCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360

QY 361 FSNKAMSDOTLRIASDGCSCVQVFTETVRRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
|||||
Db 361 FSNKAMSDOTLRIASDGCSCVQVFTETVRRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420

QY 421 GYSESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480
|||||
Db 421 GYSESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480

QY 481 AAIPA 485
|||||
Db 481 AAIPA 485

RESULT 3
Q91LD6 ID Q91LD6 PRELIMINARY; PRT; 491 AA.
AC Q91LD6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Mannitol dehydrogenase.
GN MTLDO OR PA2342.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goutry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004660; AAG05730.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENSE; 1.
KW Complete proteome.
SQ SEQUENCE 491 AA; 54350 MW; 8729A9F7E9841C39 CRC64;
```

Query Match 36.6%; Score 922; DB 16; Length 491;
Best Local Similarity 45.2%; Pred. No. 2.6e-58;
Matches 199; Conservative 66; Mismatches 171; Indels 4; Gaps 2;

```
QY 2 ITRETLKSLPANVQAPPYDIDGIRKPGIVHFGVGNFFRAHEAFYVEQILEHAPDAIVGV 59
|||||
Db 3 LNRQHLPLLATAVAPSPYDPAQIRQGIHVGFGFRAHQAAAYTDLNRRGALDWAICG 62

QY 60 VGLTGDSRSKKKAEKFAKQDCLSLTETAPSGKSTVRVWGALRDYLLAPADPEAVLKLVD 119
|||||
Db 63 AGLRSDRAMHDA--LAAQDYLTYLYELCDQPDTEVRVIGATSGMLLAEDGAELLEKLA 120

QY 120 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKPTSTVFYVVEALRRRWDAG 179
|||||
Db 121 EPAIRIVSLTITEGGYCIDDGSGEFLAEPLVRHDLANPRTPRGVGFGLCEALRRRRDTG 180
```

```
QY 180 GKAFVTMSCDNLRHNGNVARKAFLOYAKARDELAKWIEENATFNGMVDRTPTVSAEI 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 VPAFTVMSCDNLPHNGEVARAKALLAFAERLDPLGLARWIATHVSFPNAMYDRITPTMSPAH 240

QY 240 AKKLNAAAGLDDDLPLVAEDFHQWLEOFADGRPPLEKAGVQMGVDYTDWEYVIRMLN 299
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 241 RQLAQRHDEVDAMPVCEPFVQWVLEDRFSAGRPAAWKVGQFTDDVTPYEEMKIGLLN 300

QY 300 AGHVMLCFPGILVGYENVDDDAIEDSELGNLKNYLNKDVPTLKAPSGMTLEGYRDSVTS 359
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 301 GSHLALTYLGLFGRYRVFVHETLGDPLLRVYVRAFDWRDVAPLLAPVPGIDLERYKDSLVE 360

QY 360 RFSNKAMSDQTLRIASDCGSKVQVFTVTRRAIEDKRLDSRIAFGIASYLEMLRGRDEK 419
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 361 RFANRAIADQLERVCSGSSKPKFTVPTANRLIAAGRPLERVALVAAWALYLGVDHEH 420

QY 420 GGTYESSEPTYGDAEWKLAK 439
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 421 GERYPIPDPRAAECQALVAE 440

RESULT 4
O08355
ID O08355 PRELIMINARY: PRT: 493 AA.
AC O08355;
DT 01-JUL-1997 (TremBLrel. 04, Created)
DT 01-JUL-1997 (TremBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE Mannitol 2-dehydrogenase (EC 1.1.1.67) (MDH).
GN MTLd.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND SEQUENCE OF 1-20.
RC STRAIN=DSM 50106;
RX MEDLINE=97236441; PubMed=9116029;
RA Bruenker P., Altenbuchner J., Kulbe K.D., Mattes R.;
RT "Cloning, nucleotide sequence and expression of a mannitol
RT dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in
RT Escherichia coli.";
RL Biochim. Biophys. Acta 1351:157-167(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 50106;
RX MEDLINE=98121321; PubMed=9461423;
RA Bruenker P., Altenbuchner J., Mattes R.;
RT "Structure and function of the genes involved in mannitol, arabinol
RT and glucitol utilization from Pseudomonas fluorescens DSM50106.";
RL Gene 206:117-126(1998).
CC -!- FUNCTION: A MANNITOL DEHYDROGENASE WITH A BROAD SUBSTRATE
CC SPECIFICITY. SUBSTRATES INCLUDE MANNITOL, ARABITOL AND SORBITOL,
CC THESE ARE OXIDIZED TO GIVE THE CORRESPONDING KETO SUGARS. THIS
CC ENZYME WILL ALSO CATALYZE THE REDUCTION OF FRUCTOSE AND XYLULOSE.
CC -!- CATALYTIC ACTIVITY: D-MANNITOL + NAD(+) = D-FRUCTOSE + NADH.
CC -!- SUBUNIT: MONOMER.
CC -!- INDUCTION: BY MANNITOL, ARABITOL AND SORBITOL. NOT INDUCED BY
CC FRUCTOSE OR XYLULOSE.
CC -!- MISCELLANEOUS: THE MTLD PROTEIN IS ENCODED BY THE MTL EFGKDYZ
CC OPERON. THIS OPERON ENCODES PROTEINS FOR THE UPTAKE AND
CC UTILIZATION OF MANNITOL, ARABITOL AND SORBITOL.
CC -!- SIMILARITY: AAC04472.1;
DR EMBL: AF007800; AAC04472.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLdHGRGNASE.
DR PROSITE: PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 29 40 NAD (BY SIMILARITY).
SQ SEQUENCE 493 AA: 54497 MW; 7C12DFA443CEA443 CRC64;
```

```
Query Match 35.3%; Score 888; DB 2; Length 493;
Best Local Similarity 42.2%; Pred. No. 7.4e-56;
Matches 181; Conservative 74; Mismatches 170; Indels 4; Gaps 2;

QY 2 ITRETLKSLPANVQAPPYDIDGIKPGIVHFGVGNFRFAHEAFYVEQILE--HAPDWAIVG 59
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 3 LNKQNLTLQAPLEVKLPAYTLADTROGIAHIGVGFRHQAAYYTDALMNTGGLDWSICG 62

QY 60 VGLTCSGSKKAEFEKQACDCLYSLTETAPSGKSVRVMGALROYLLAPADPEALVKHLV 119
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 63 VGLRSEDR--KARDDLQAGDYLFTLYELGDDTDDTEVRVIGSISDMLLAEDSAQALDKLA 120

QY 120 DPAIRIVSWITTEGYNINETTGAFDLENAAYKADLKNPEKSTVFGYVVEALRRWDAG 179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 121 SPEIRIVSLTITEGCGYCIDDSNGEPMALPQIQHDLAHPSSPKTVFGFICAALTQORRAAG 180

QY 180 GKAFVTMSCDNLRHNGNVARKAFLOYAKARDELAKWIEENATFNGMVDRTPTVSAEI 239
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 181 VPAFTVMSCDNLPHNGEVARAKALLAFAERLDPLGLARWIATHVSFPNAMYDRITPTMSTAH 240

QY 240 AKKLNAAAGLDDDLPLVAEDFHQWLEOFADGRPPLEKAGVQMGVDYTDWEYVIRMLN 299
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 241 RQLHDEHGIDDAMPVCEPFVQWVLEDRFVNGRPAAWKVGQFTDDVTPYEEMKIGLLN 300

QY 300 AGHVMLCFPGILVGYENVDDDAIEDSELGNLKNYLNKDVPTLKAPSGMTLEGYRDSVTS 359
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 301 GSHLALTYLGLFGRYRVFVHETMNDPLFVAYMAYMDLDTNPLAPVPGIDLTQYKOTLVD 360

QY 360 RFSNKAMSDQTLRIASDCGSKVQVFTVTRRAIEDKRLDSRIAFGIASYLEMLRGRDEK 419
      ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||: ||||:
Db 361 RFSNOAIADQLERVCSGSSKPKFTVPTINRLIADGRETERAALVVAWALYLGVDEN 420

QY 420 GGTYESSEP 428
      ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 421 GVSYPIDDP 429

RESULT 5
O98D06
ID O98D06 PRELIMINARY: PRT: 492 AA.
AC O98D06;
DT 01-OCT-2001 (TremBLrel. 18, Created)
DT 01-OCT-2001 (TremBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)
DE Mannitol dehydrogenase.
GN MLL4920.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003005; BAB51465.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLdHGRGNASE.
DR PROSITE: PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 492 AA: 53518 MW; 4D20E1136C3C1373 CRC64;
```

```
Query Match 31.7%; Score 797; DB 16; Length 492;
Best Local Similarity 39.3%; Pred. No. 2.8e-49;
Matches 174; Conservative 72; Mismatches 187; Indels 10; Gaps 6;
```

Qy	62	LTGSDRSKKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKHVLVDP	121
Db	63	L-----RSPETRDALHPDQGLYTLVAVODGEG-SELSVVGSVELLCPADDPEAVLRMDAP	117
Qy	122	AIRIVSMTITTEGGYNTINETGAPDLENAAVKADLKNPEKSPSTVFVGVVEALRRRRWDAGGK	181
Db	118	GTRIVLSLTITEKGYCHNPATGTTLDEGHDPDVLVDLANPARPSAIGFIVEAISRRYSAGIA	177
Qy	182	AFVMSCDNLRHNGNVARKAFGLYAKAROBELAKWIEENATFPNGMVDRIITPTVSAEITAK	241
Db	178	PFTLLSCDNLPGNGHVLKRLVTOFAEAROPALAA-VVRNVASPTMVDRIVPATTDSDRS	236
Qy	242	KLNAASGLDDDLPLVAEDFHQWVLEDOQADGRPPLEKAGVQMVGVDYDWEYVKIRMLNAG	301
Db	237	AVASAMGLEDAWPIETPEFRQWVIEEDFPLGRPAWEKACALKFQDVSAFEFKMLRLNGS	296
Qy	302	HVMLCFPGIILVGVENVDDAIEDSELGNLKNLYLNKDVIPTLKAPSCMTLEGYRDSVISRF	361
Db	297	HSTLVLGYLAGAETVADAMALAGMEALVEGLMRHVSSTLPPELPGFDLPAYRAELLQRF	356
Qy	362	SNKAMSDOTLRIASDGCSKVQVFWTETVRRATEDKRDLSRIAFGIASYLEMLRGDEKGG	421
Db	357	RNPALRHRTWQIAMDGSQKLPQRLLLGISIRDRLOAGAGYDRLAGVAAWMRYARGLDEAGR	416
Qy	422	TYESSEP	428
Db	417	PIDVRDP	423
RESULT	7		
ID	08X516		
		PRELIMINARY:	
		PRT:	486 AA.

01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DT Mannonate oxidoreductase (D-mannonate oxidoreductase).
 GN UYXU OR Z5921 OR ECS5282.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=83334;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RN Nature 409:529-533(2001).
 RP [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Omishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Iida C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Hata T., Takami H., Honda T., Sasakiawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RN DNA Res. 8:11-22(2001).
 RL ENBL; AE005663; AAC59505.1; -;
 DR ENBL; AP002569; BAB38705.1; -;
 DR InterPro; IPR000669; Mannitol_dh.
 DR Pfam; PF01232; Mannitol_dh; 1
 DR PRINTS; PR00084; MTL0DHRCNASE.
 DR PROSITE; PS00974; MANNITOL_DHGHEINASE; 1.
 SQ Complete proteome.
 KW SEQUENCE 486 AA; 53656 MW; C519DD28763A41DC CRC64;

KL ENBL RES : 8; LI = 22 (2001) ;
 DR ENBL; AE005663; AAG59505.1; -;
 DR ENBL; AP002569; BAB38705.1; -;
 DR InterPro: IPR000669; Mannitol_dh.
 DR Pfam: PF01232; Mannitol_dh; 1
 DR PRINTS: PR00084; MTL0HDRGNASE
 DR PROSITE; PS00974; MANNITOL_DHGENSE; 1.
 DR Complete proteome.
 KW SQ SEQUENCE 486 AA; 53656 MW; C519DD28763A1DC CRC64;

	Query Match	30.8%;	Score	775.5;	DB	16;	Length	486;
	Best Local Similarity	36.4%;	Pred. No.	9.9e-48;				
	Matches	177;	Conservative	87;	Mismatches	205;	Indels	17;
	Gaps							
Qy	9	SLPANVOAPPYDIDGIRKPIGVHFGVGNFFRAHEAFYVEQIILEHA-PDWAIWVGVL-TGSD	66					
Db	8	NLP--VARPSWDHSRLSRIIVHLGGCAFHRAHQALYTHHLLSTDSWGICEVNLMPGND	65					
Qy	67	RSKKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALRDYLLAPADP-EAVLKHLDVPATRI	125					
Db	66	R-V-LVIENLKQQQLLYTVAEKG-AESTELKTIGSMKEALHPEDIDCEGILNAMAARPQTAI	122					
Qy	126	VSMITTEGGYNINETGAFDLENAVAADLANPEKSPVFYGVVEALRRRDDACGKAFTV	185					
Db	123	VSLVTKEGYCADAAAGOLDLNNPLIKHDLNPAPTAKSAIGTIIVEALKRREKGLUKAFTV	182					
Qy	186	MSCDNLRHNGNVARKAFLGYAKARDELAKWIEENATIPNGMVDRIPTVSAEIAKKLNA	245					
Db	183	MSCDNVBRENGHVAKVAVLGQAARDPOLAAIEENVTFPCTMVDRIVPAATPETLQEIAD	242					
Qy	246	ASGLDDDLLPLVAEDFHWLVLEDQFADGRPPLEKAGVQMVGDTVWEYVKIRMLNAGHVML	305					
Db	243	QLGYDPCAACEPFROWIIEDNFVNRPDKVGAOFVADVFPFMKMLRMJNSHSL	302					
Qy	306	CFFGILGYEVNVDDAIEDSELLGNLNKYLNKDVIPTIKAPSGMTLEGYRDSVISRFNSKA	365					
Db	303	AYLGVLGYETIADTMTPDYRKAAFALMWGEQAPTLSMPEGTDLNAYATLLIERFSNPS	362					
Qy	366	MSDOTLIARIADGCSKVQVFWTTETVRRALDGSRIFAEGIASYLEMLRGDEKGGTYES	425					
Db	363	LRHRTWIAMDGSOKLPQRLLDPVRLHLONGSSWRHLALGAVGMRYVTQGVDGECGNADIV	422					
Qy	426	SEPTYGDAEWKLAKADDFESSLKLPAPDGMWRDLDTSELDOK-----VIVLRKIIREKGV	479					
Db	423	VDPMPL--AEFQKINA-QYQADRVKALLGLSGIFADDLPQNADFVGAVTAAYQQLCERGA	479					
Qy	480	KAAIPA	485					
Db	480	RECVA	485					

```

RESULT 8
Q8X506
ID Q8X506 PRELIMINARY; PRT; 486 AA.
AC Q8X506;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative oxidoreductase.
GN YDFI_OR_Z2155 OR ECS2151.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RX [1]
SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / EDL933 / AFCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[2]
SEQUENCE FROM N.A.
RP STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

```

RA	Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.
RA	Kunhara S., Shiba T., Hattori M., Shinagawa H.
RT	"Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i>
RT	O157:H7 and genomic comparison with a laboratory strain K-12.";
RL	DA Res. 8:11-22(2001).
DR	ENBL; AE005350; BAG56220.1; -
DR	ENBL; AP002557; BAB35574.1; -
DR	InterPro; IPR000669; Mannitol_dh.
DR	Fram; PF01232; Mannitol_dh; 1.
DR	PRINTS; PR00084; MTLDRHGNASE.
DR	PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW	Complete proteome.
SQ	SEQUENCE 486 AA; 53581 MW; 1AFD234D83108D9E CRC64;

Query Match	30.6%;	Score	769.5;	DB	16;	Length	486;
Best Local Similarity	37.9%;	Pred. No.	2.7e-47;				
Matches	184;	Conservative	78;	Mismatches	201;	Indels	23;
Gaps							

Qy	10	LPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYFEQI--LEHAPDWAIVTGVGLTGS	688
Db	6	LSAKATLPVYDRNNLAPRIVHLGFGAFHRAHQGVADILATEHFSDMGYEYVNLIGGE--	63
Qy	69	KKAAEEFKADCLYSLETETAPSKSTVRYVNGALRDYLLAPADP-EAVLKHLDVPATRIYS	127
Db	64	-QQIADLHOODNLYTVAEMS-ADAMTARVYGVYKALHVQIDGLETVLAAACEPQIAIYS	121
Qy	128	MTITEGGYNIETTTGAFDLENAVAKADLKNPEKPTVFGYVVEALRRRDAGCKAFTVMS	187
Db	122	LTITEKGYFHSIPATGQLMDHPVADVQNPQPKTATGIVIVEALARRKAAGLPAFTVMS	181
Qy	188	CNDLRHNGNVARKAFLGYAKARPELAKWITEENATPFGMVDRITPTVSAEIAKKLNAAS	247
Db	182	CNDNPENGVHVRDVTYSQAQADVKKLAQITEDNVTFTSTVMDRIVPAVTEDTLAKIEQLT	241
Qy	248	GLDDDLPLVAEDFHOWVLEDFQADGRPPLEKAGQVMQGVDTDHEYVKIRMLNAGHVMLCF	307
Db	242	GVRDPAGVACEPFRQWIEDNFVAGRPEWEKAGELVSDVLPEEMKRLNGSHSFLAY	301
Qy	308	PGILVGYENYDDATIDSELGLNKLNYLNKDVIPTLKPAGSMTLLEGYRDSVISRFSNKAMS	367
Db	302	LGYLAGYOHINDCMEDEHYRHAAYALMLQEQAPTLKV-QGVLDQDYANRLIARYSNPALR	360
Qy	368	DOTLRASDCSKVOVFWETVR-RAIEDKRLSRIFAAGIASYLEMLRGRDEKGGTYES	425
Db	361	HRTWQIAMDGSQKLPQRLMDSVRWHLAHSKFDL--LALGVAGMWRVYGVGVDEQGNPIEI	418
Qy	426	SEP-----TVGDAEWKLAADDFESSLUKLPFAFGWRDLDTSELDQKIVLVRKLTIREK	477
Db	419	SDPLLPIQKAVOSSAEGKAR----VQSLLAIKAFGGDLPNLSLFAKVTAYLSLLAH	474
Qy	478	GVKRAAI 483	
Db	475	GAKATV 480	

RESULT	9	
O68111		
ID	O68111	PRELIMINARY; PRT; 478 AA.
AC	O68111;	
DT	01-AUG-1998 (TrEMBLrel. 07, Created)	
DT	01-AUG-1998 (TrEMBLrel. 07, Last sequence update)	
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)	
DE	Mannitol 2-dehydrogenase (EC 1.1.1.67).	
OS	Rhodobacter capsulatus (Rhodospseudomonas capsulata).	
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;	
OC	Rhodobacter.	
OX	NCBI_TaxID=1061;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	STRAIN=SB1003;	
RX	MEDLINE=97404404; PubMed=9256491;	
RT	Vlcek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;	
RT	"Sequence of a 189-kb segment of the chromosome of Rhodospirillum rubrum"	

RT	capsulatus SB1003. ";
DR	Proc. Natl. Acad. Sci. U.S.A. 94: 9384-9388 (1997).
RL	EMBL; AF010496; AAC16201.1; -.
DR	InterPro: IPR000669; Mannitol_dh.
DR	Pfam: PF01232; Mannitol_dh; 1.
DR	Pfam: PR00084; TMDHGRNASE.
DR	PROSITE; PS00974; MNLDRHGRNASE; 1.
DR	Oxidoreductase.
KW	SEQUENCE 478 AA: 51442 MW: 8427F89B6F2A4F38 C

Query Match 30.5%; Score.768; DB 2; Length 478;
Best Local Similarity 36.4%; Pred. No. 3.3e-47;
Matches 176; Conservative 84; Mismatches 202; Indels 22; Gaps 9;

Qy	10	LPANQAPPYDIDGIPKGVHFCVGNFFRAHEAFVYEQI--LEHAPDMAIVGVGLTGSDR	67
Db	1	MPYVAAAPRYDRAQLTPGVHIGLGNFRAHQAVYLDLFLGEGHDWALICAGVRAPDA	60
Qy	68	SKKAKEEFAQOCLYSITETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRVS	127
Db	61	AMR--EVLLAQDGLSSVIELAP-GAHSARVIGAMIDFVAVQPCNAALIAMRPELRVS	117
Qy	128	MTTEGGYNINETGAFDLENAAVKADLKNPEKPSYFVGVVEALRRRWADGCAFTVMS	187
Db	118	LTVTEGGYYIDPNTGTFSPDHPDIQADARPAPDPTAFGAIITAAALKMRRAAGVAPFTVMC	177
Qy	188	CNLRNGNVARAKFYGLKARPELAKWTEENATFPGNMVDRITPTVSAETAKKLNAAS	247
Db	178	CNDVPHNGHVTRDVAVGLARFLDAELADWTHASVAFPSNMVDRITP-ATGDRERAMAADL	236
Qy	248	GLDDDLPLVAEDFQHWVLEDOQFADGRPPLEKAGCVQMVGVDTWNEYVKYIRMLNAGHVMLCF	307
Db	237	GLDDAAPVTCETFRQWIEDKFPAGRPALKEKGVGVTFTPHWDKFETMKIRILNGGHAIAY	296
Qy	308	PGILVGVENVDDAIEDSELLGNLKNLYNKDVIPTLKA PSGMTLEGYRDSVIRSFNKAMS	367
Db	297	PGGLADIATFHEAMADQIRAFNLNRVLTAEVLVIPPVPGQDLNDYKALIIIFRFSNPEVA	356
Qy	368	DOTLRIASDGCCKVQVFWTETVRRRIEDKRDJSRIAGFIASYLEMLURKDEKGGTYESSE	427
Db	357	DTVRRRLCDGNSRQPKFIIPSIRDRIATGGSFGDLVLLSALWCYRCFGTTEGSAEIAPN	416
Qy	428	PTYGDAEMKLAKADDRESSLKLPAFDGWRDLDT--SELDOQKVLVLR-----KIIIREKGV	479
Db	417	PNW----ERLVAVA---QAAKDRFA--AWLEMDIYAEVGNRPAVIESFTTALKAIWARGS	468
Qy	480	KAAI 483	
Db	469	RAVI 472	

RESULT	10
Q92N07	
ID	PRELIMINARY; PRRT; 494 AA.
AC	Q92N07;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DC	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE	Probable mannitol 2-dehydrogenase protein (EC I.I.1.67).
GN	MTLK OR R02440 OR SMC01501.
OS	Rhizobium meliloti (Sinorhizobium meliloti).
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC	Rhizobiaceae; Sinorhizobium.
OX	NCBI_TaxID=382;
RP	[1]
RN	SEQUENCE FROM N.A.
RC	STRAIN=1021;
RX	MEDLINE=21396507; PubMed=11481430;
RA	Capella D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA	Boistard P., Becker A., Boudry M., Cadieu E., Dreano S., Gloux S.,
RA	Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA	Pohl T., Portetelle D., Puehler A., Furnelle B., Ramsperger U.,
RA	Renard C., Thebaud P., Weidner M., Weidner S., Galibert F.;

RT "Analysis of the chromosome sequence of the legume symbiont
RT *Sinorhizobium meliloti* strain 1021.",
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591790; CAC47019.1; -
DR Interpro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PROSITE; PS00974; MANNITOL_DHENASE; UNKNOWN_1.
DR Oxidoreductase; Complete proteome.
KW SEQUENCE. 494 AA; 53822 MW; 7533287D8EF58F24; CRC64;

Query Match 30.2%; Score 760; DB 16; Length 494;
Best Local Similarity 36.1%; Pred. No. 1.3e-46;
Matches 176; Conservative 83; Mismatches 208; Indels 20; Gaps

Qy	6	TLKSLPANWOAPPDYIDIGTKPGTVHFGVGNFPAHAEFAFVEQI--LEHAPDWAIVCVGLT	63
Db	9	TLDAVKATAGVPNGRHLRAGTVHFGVGNFRAHQAQVLDLDFNLGRDWDMAIIAGAVL	68
Qy	64	GSDRSKKAEFEQAQDCLYSLTETAPSGKSTVVMGALRDYLLIAPADPBAVLKHLVDPAI	123
Db	69	PSD--KVMRKLEAQDFLTTVVE-QDNRTGAHVTGAMTAY-LEPGDTPAIVAOQLASPLI	124
Qy	124	RIVSMTTIEGGYINETTGAFDLENAAVRADLKNPEKPSTVFGYVVEALRRRWDAGGKAF	183
Db	125	RIVSLATTIEGGYFIDPASGVFDEPAHFAIVEDARDPAAPKTVFGLIUGLAERRAKGIPPF	184
Qy	184	TVASCDRLRNGNVARKAFUGYAKAKDPDELAKKIEENATFPNGMVDRIPTTVSAETAKKL	243
Db	185	TIMSCDNI PCNGEVTHAAVSGLARLSDPCFADWDIDANVAFPMGMDRIITPATGAREIGV	244
Qy	244	NASAGLDDLLPLVAEPHQWLQEDQADGRPPLEKAGVQMGVDVTDWEYVKIRMLNAGHV	303
Db	245	ASQYIGDAMPVFECEFKQWLVEDRFPPQPALEEYGVQFVDPVAPYEHMKIRLINGGHA	304
Qy	304	MLCFPGTLVGYENVDDAIEDSELGLNKNYLNRKDVIPTLKAPEGMTLEGYRDSVTSRFSN	363
Db	305	AIAPPAALLDIHFVHEAMEPLITRAFLSKLEHDEIIPVPPVPTDOLKOYKLIETRFNSN	364
Qy	364	KAMSDOTLRIASDGCSSKVQVFTWTVRRATIEDKRDLGSRTAFGTSYLEMLRGRDEKGGTY	423
Db	365	PKIGDTVARLAQAQDGSNRQPKFILPSTADRLRRGEDVVGLSVLSALWCRYFAGKSDSG--	421
Qy	424	ESSEPTYGDAE-----WKLAKADDDFESSLKLPAFGWRDLDSTLEQDKQIV-LRKIIRE	476
Db	422	--KEIVENDANADLHAAVAAKDDPMAFLALS--DIFGDVAQSDLFRRRFAHALKVLWE	477
Qy	477	KGKAAI	483
Db	478	KGTRATL	484

RESULT	11
Q826Z6	
ID	Q8Z6Z6 PRELIMINARY; PRT; 488 AA.
AC	Q8Z6Z6;
DT	01-MAR-2002 (TrEMBLrel. 20, Created)
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE	Putative D-mannosate oxidoreductase (EC I.I.1.57).
GN	STY1553.
OS	Salmonella typhi.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Salmonella.
OX	NCBI_taxid=601;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CT18;
RC	MEDLINE=21534947; PubMed=11677608;
RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,

```
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL627270; CAD01805.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRHGNASE.
DR PROSITE; PS00974; MANNITOL_DHEGENASE; 1.
KW Oxidoreductase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 488 AA; 53922 MW; 18F64DE2ACCA7781 CRC64;

Query Match          30.2%; Score 759.5; DB 16; Length 488;
Best Local Similarity 36.3%; Pred. No. 1.4e-46;
Matches 175; Conservative 87; Mismatches 205; Indels 15; Gaps 7;

QY 10 LPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFVVEQI-LEHAPDWAIVGVLGSGRS 68
   | | | | | : : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 6 LSAKATLPDYDRAALAARMVHLGFCFAHRAHQGVYTDILAAEQHSDWGYYEVLNIGGE-- 63
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 69 KKAEEFAQDCCLYSLETETAPSGKSTVRVMGALRDYLLAPADP-EAVLKHLVDPALRIVS 127
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 64 -QOIADLKQODNLYTVAEMSAEA-WTARVGVVKAALHVQVGDGLERVLAAACEPQIAIVS 121
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 128 MTITEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYGVVVEALRRRDWAGGKAFTVMS 187
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 122 LTITEGYCHSPATGQLLEHPMIAADLQNPQHPITAPGIIIVEALARRKAAGLPAFTVMS 181
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 188 CNLHNGNVARKAFGLYAKARDEPDLAKWIEENATFPNGWVDRIPTVSAETAKKLNAA 247
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 182 CNMPENGHVTRQVVTAYAREVDAELAIWIEQNVTFPSTWMDRIVPAVTPETLDKIEQLT 241
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 248 GLDDDLPLVAEDFHOWVLEDFADGRPPLEKAGVOMGVDVTDWEYVKTRMLNAGHVMLCF 307
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 242 GVRDPAGVACEFFROWVIEDTVAGRPQWENAGATLVADVVPFEEMKRLMNGSHSFLAY 301
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 308 PGILVYENVDVAIEDSELLGNLKNYLNKDVITPTLKAPSGMTLEGYRDSVSRFSNKAMS 367
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 302 LGYLAGYQHINDCDDNNYRLTAQALMLREQAPTLLK-VGVDLQRYADQILIARYNPALR 360
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 368 DQTLRIASDGSCKVQVFWTETVRRRAIEDKRLSRITAFGIASYLEMLRGRDEKGGTYESSE 427
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 361 HRTWQIAMDGSQKLPQRLMDSVVRHLNHSDFDLALGVAGWMRYVGVGVDEGKAIDVSD 420
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 428 PTYGADEWKLAKADDFESSKLPAFDGWRDLDTSELDQKVIIVLRKI-----IREKGVKA 481
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 421 PLLPVIQRAVANSEEGASRVK--ALLGMAEIFGNLPOAARFTQKVQAEYDSLITYGAKA 478
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 482 AI 483
   : :
Db 479 SV 480

RESULT 12
Q8ZPI8 PRELIMINARY; PRT; 488 AA.
AC Q8ZPI8;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE Putative mannitol dehydrogenase.
GN YDFI OR STM1508.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
```

```
RA Courtney L., Potwolik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
DR EMBL; AE008765; AAL20427.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRHGNASE.
DR PROSITE; PS00974; MANNITOL_DHEGENASE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 488 AA; 53896 MW; 704BF8C27F4407A6 CRC64;

Query Match          30.0%; Score 755.5; DB 16; Length 488;
Best Local Similarity 36.3%; Pred. No. 2.8e-46;
Matches 175; Conservative 86; Mismatches 206; Indels 15; Gaps 7;

QY 10 LPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFVVEQI-LEHAPDWAIVGVLGSGRS 68
   | | | | | : : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 6 LSAKATLPDYDRAALAARMVHLGFCFAHRAHQGVYTDILAAEQHSDWGYYEVLNIGGE-- 63
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 69 KKAEEFAQDCCLYSLETETAPSGKSTVRVMGALRDYLLAPADP-EAVLKHLVDPALRIVS 127
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 64 -QOIADLKQODNLYTVAEMSAEA-WTARVGVVKAALHVQVGDGLERVLAAACEPQIAIVS 121
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 128 MTITEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYGVVVEALRRRDWAGGKAFTVMS 187
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 122 LTITEGYCHSPATGQLLEHPMIAADLQNPQHPITAPGIIIVEALARRKAAGLPAFTVMS 181
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 188 CNLHNGNVARKAFGLYAKARDEPDLAKWIEENATFPNGWVDRIPTVSAETAKKLNAA 247
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 182 CNMPENGHVTRQVVTAYAREVDAELAIWIEQNVTFPSTWMDRIVPAVTPETLDKIEQLT 241
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 248 GLDDDLPLVAEDFHOWVLEDFADGRPPLEKAGVOMGVDVTDWEYVKTRMLNAGHVMLCF 307
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 242 GVRDPAGVACEFFROWVIEDTVAGRPQWENAGATLVADVVPFEEMKRLMNGSHSFLAY 301
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 308 PGILVYENVDVAIEDSELLGNLKNYLNKDVITPTLKAPSGMTLEGYRDSVSRFSNKAMS 367
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 302 LGYLAGYQHINDCDDNNYRLTAQALMLREQAPTLLK-VGVDLQRYADQILIARYNPALR 360
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 368 DQTLRIASDGSCKVQVFWTETVRRRAIEDKRLSRITAFGIASYLEMLRGRDEKGGTYESSE 427
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 361 HRTWQIAMDGSQKLPQRLMDSVVRHLNHSDFDLALGVAGWMRYVGVGVDEGKAIDVSD 420
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 428 PTYGADEWKLAKADDFESSKLPAFDGWRDLDTSELDQKVIIVLRKI-----IREKGVKA 481
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
Db 421 PLLPVIQRAVANSEEGASRVK--ALLGMAEIFGNLPOAARFTQKVQAEYDSLITYGAKA 478
   : : | | | | | : : | | | | | : : | | | | | : : | | | | | : : | | | | |
QY 482 AI 483
   : :
Db 479 SV 480

RESULT 13
Q8ZGK9 PRELIMINARY; PRT; 490 AA.
AC Q8ZGK9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DE Putative D-mannanate oxidoreductase.
GN YP01280.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
```


OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSCI412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AF008844; AAL22010.1; -;
DR EMBL; AL627277; CAD02968.1; -;
DR InterPro; IPR000669; Mannitol_dh.
DR InterPro; IPR001993; Mitoch_carrier.
DR Pfam; PF01232; Mannitol_dh; 1;
DR PRINTS; PR00084; MFLDHRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
KW Oxidoreductase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 490 AA; 54007 MW; A521610BB383A9FF CRC64;

Query Match 29.0%; Score 730; DB 16; Length 490;
Best Local Similarity 36.5%; Pred. No. 1.9e-44;
Matches 156; Conservative 83; Mismatches 182; Indels 6; Gaps 5;

QY 5 ETLKSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHA-PDWAIVGVGL- 62
DB 3 QNIATAQVSVARPNWDKSLVSRIVHLGCGAFRAHQALFTHHLEKSDSDMGICEVNL 62
QY 63 TGSRSKKKAEFEKQDCILYSLETAPSGKSTVRVMGALRDYLLAPADPEA-VLKHLVDP 121
DB 63 PGND--ARLIANKAQNLYTVARG-AESTELKIIGSMKEALHPEFDGHGILAAMARP 119
QY 122 ATRIVSWTTTEGYNINETTGAFDLENAAVKADLNKPEKSPVFGYVVYVVALRRRWDA 181
DB 120 ETATVSLVTEKGYCTDPASGELDVNNPLIQNDLAHPQPKSAIGYIVEALNMRREQ 179
QY 182 AFTVNSCDNLRLNGVNVKAFGLYKAKARPELAKEENATPPNGMVDRTPTVSAETAK 241
DB 180 AFTVLSCDNVRENGHVAKAANVGLAKARDAALAAWIAADNVTFCTWVDRIVPAATEET 239
QY 242 KLNAAAGLDDDLPLVAEDFHOWLEDQFADGRPPLEKAGVQMVGVDTWYEVKIRMLNAG 301
DB 240 LVADQLGVYDPCAIACEPPFQWVIEDNFVNGRPDMDTVGAQFVADVVPFEMKRLNCS 299
QY 302 HVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAPSGMTLEGYRDSVSRF 361
DB 300 HSFLAYLGYLGGYDTADTMTNPAYRRAALALMDLDEQAPTLSPMEGTDLGYVANLLIARF 359
QY 362 SNKAMSDOTLRASDGSCKSVQVFTWETVPRATEDKRDLSRIAFTASYLEMLRGDEKGG 421
DB 360 TNPULKHTWQIAMDGSQKLQRLDLPVRLHLQGGDDYRRLTLGVAGWMRYVGGVDEQCK 419

QY 422 TYESSEP 428
DB 420 TIDVVDP 426
Search completed: May 8, 2003, 16:33:38
Job time : 41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 19, 2003, 00:18:20 ; Search time 10357 Seconds
(without alignments)
11563.005 Million cell updates/sec
Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttgatgcctgcaggt.....tccggcgccctgaagctt 4115
Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0
Searched: 2054640 seqs, 14551402878 residues
Word size : 0
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4073	99.0	4073	1	AB028937	AB028937 Gluconoba
2	40	1.0	4643	12	AF027127	AF027127 Eukaryoti
3	38	0.9	91	12	SYNIIFOKI	M30778 MboII/FOKI
4	38	0.9	360	6	AX256425	AX256425 Sequence
5	38	0.9	416	5	DRE238981	AJ238981 Danio rer
6	38	0.9	1195	6	I09132	I09132 Sequence 1
7	38	0.9	2442	1	MTGLMS	AJ000333 Mycobacte
8	38	0.9	4404	12	SYNCATBLB	M80483 Cloning vec
9	38	0.9	4496	12	BLCAT2DNA	X64410 plasmid pBL
10	38	0.9	4711	12	AF027128	AF027128 Eukaryoti
11	38	0.9	7055	12	AF102233	AF102233 Transposo
12	38	0.9	8142	12	AF076213	AF076213 Promoter
13	38	0.9	10737	12	XXU02428	U02428 Cloning vec
14	38	0.9	10850	12	U02455	U02455 Cloning vec
15	37	0.9	92	6	A25917	A25917 Expression
16	37	0.9	92	6	A25918	A25918 Expression
17	37	0.9	341	11	G49583	G49583 SHGC-60412
18	37	0.9	586	6	AR007549	AR007549 Sequence
19	37	0.9	586	6	AR029479	AR029479 Sequence
20	37	0.9	586	6	I90252	I90252 Sequence 10
21	37	0.9	916	11	G63570	G63570 SHGC-141568
22	37	0.9	2907	12	XXU13871	U13871 pT73D clon
23	37	0.9	3796	6	E49323	E49323 Infectious
24	37	0.9	4674	12	XXU02437	U02437 Cloning vec
25	37	0.9	4910	6	AR027100	AR027100 Sequence
26	37	0.9	4910	6	AR036062	AR036062 Sequence
27	37	0.9	4910	6	AR160052	AR160052 Sequence
28	37	0.9	4910	6	AR209764	AR209764 Sequence
29	37	0.9	5027	12	AF276982	AF276982 Integrati
30	37	0.9	5321	12	AF129072	AF129072 Cloning v
31	37	0.9	6282	12	AF069723	AF069723 Expressio
32	37	0.9	6534	12	AF069724	AF069724 Expressio
33	37	0.9	6541	12	XXU02433	U02433 Cloning vec
34	37	0.9	6894	12	XXU02435	U02435 Cloning vec
35	37	0.9	7036	12	AF069722	AF069722 Expressio
36	37	0.9	7069	12	U02442	U02442 Cloning vec
37	37	0.9	7164	6	AX195206	AX195206 Sequence
38	37	0.9	7164	12	U02451	U02451 Cloning vec
39	37	0.9	7481	12	XXU02438	U02438 Cloning vec
40	37	0.9	7522	12	AF069717	AF069717 Expressio
41	37	0.9	7612	6	AX009487	AX009487 Sequence
42	37	0.9	7897	6	AR093383	AR093383 Sequence
43	37	0.9	7897	6	AR142583	AR142583 Sequence
44	37	0.9	8540	6	AR198722	AR198722 Sequence
45	37	0.9	8932	6	AR069417	AR069417 Sequence

ALIGNMENTS

RESULT 1
AB028937
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

AB028937
Gluconobacter oxydans gene for sorbitol dehydrogenase, complete
cgs.
AB028937
AB028937.1 GI:9049448
sorbitol dehydrogenase.
Gluconobacter oxydans (strain:G624) DNA.
Gluconobacter oxydans
Bacteria; Proteobacteria; alpha subdivision: Acetobacteraceae;
Gluconobacter.
1 (sites)
Shibata,T., Ichikawa,C., Matsuura,M., Takata,Y., Noguchi,Y.,
Saito,Y. and Yamashita,M.

QY 1657 CTAGCGATGGGTGTTTCCAAAGTTTCAGGTGTTCTTGGACGAAACCGTGGTCCGGCGATCG 1716
DB 1621 CTAGCGATGGGTGTTTCCAAAGTTTCAGGTGTTCTTGGACGAAACCGTGGTCCGGCGATCG 1680
QY 1717 AAGACAAGCGGAGCTGTACAGTATAGGTTTCGGAATTCATCTCAAAATGCTGC 1776
DB 1681 AAGACAAGCGGAGCTGTACAGTATAGGTTTCGGAATTCATCTCAAAATGCTGC 1740
QY 1777 GTGGTCGCGACGAGAAGCGGGAGCTATGAATCGTCCGAGCCGACTTATGGCGACGCG 1836
DB 1741 GTGGTCGCGACGAGAAGCGGGAGCTATGAATCGTCCGAGCCGACTTATGGCGAGCGCG 1800
QY 1837 AATGGAAGTTGGCAAGCGGAGCGAGCTTCGAAAGCTCTCTGAAGCTCCCGCGCTTCGATG 1896
DB 1801 AATGGAAGTTGGCAAGCGGAGCGAGCTTCGAAAGCTCTCTGAAGCTCCCGCGCTTCGATG 1860
QY 1897 GTGGCGCGATCTGGATACGTCCGAACCTGGATCAAAAGTTCATCTGCTCGGGAAGATCA 1956
DB 1861 GTGGCGCGATCTGGATACGTCCGAACCTGGATCAAAAGTTCATCTGCTCGGGAAGATCA 1920
QY 1957 TCCGCGAAAAGGGGTAAAGCCGCAATCCCGGCTCGAATTCGGCTTTTAGGGTAGCGAC 2016
DB 1921 TCCGCGAAAAGGGGTAAAGCCGCAATCCCGGCTCGAATTCGGCTTTTAGGGTAGCGAC 1980
QY 2017 TGAACAGAAAACCGCGCTCTGGAAGGAGCGCGGTTTTTTTTATGCTCAGATCTGTCCCA 2076
DB 1981 TGAACAGAAAACCGCGCTCTGGAAGGAGCGCGGTTTTTTTTATGCTCAGATCTGTCCCA 2040
QY 2077 TCAGGACAAGGATCAGCAGACACGATCAGGACAGTCCGCTCGAGGGGAGCCCAAT 2136
DB 2041 TCAGGACAAGGATCAGCAGACACGATCAGGACAGTCCGCTCGAGGGGAGCCCAAT 2100
QY 2137 TCGAACGTGACGGCATGACGGCAGCGACCGAGATCAGGATTAACAAGAGGATCAGTCC 2196
DB 2101 TCGAACGTGACGGCATGACGGCAGCGACCGAGATCAGGATTAACAAGAGGATCAGTCC 2160
QY 2197 CATGGCACATCTCTTGGCCGGTTGAGACTGGTCTGTGTTCCGGGTGTCTAAAAAGTTTC 2256
DB 2161 CATGGCACATCTCTTGGCCGGTTGAGACTGGTCTGTGTTCCGGGTGTCTAAAAAGTTTC 2220
QY 2257 CGTAGGGCGGGAAGATCAAGCTGTGCGTTCGCGCTTAATCCGCTCCCAAGCGCATTG 2316
DB 2221 CGTAGGGCGGGAAGATCAAGCTGTGCGTTCGCGCTTAATCCGCTCCCAAGCGCATTG 2280
QY 2317 ATCGGGGCCACCCGGTCTGTGCGGTTGCGCTCTGTCTGACATAGGTTTCTGGGCC 2376
DB 2281 ATCGGGGCCACCCGGTCTGTGCGGTTGCGCTCTGTCTGACATAGGTTTCTGGGCC 2340
QY 2377 AGCAGTCCGGATGATGTTCCGGATCAGGTTGCGGACGCGACGCGGATTTCTGTGTCA 2436
DB 2341 AGCAGTCCGGATGATGTTCCGGATCAGGTTGCGGACGCGACGCGGATTTCTGTGTCA 2400
QY 2437 GTTCGGCTGGGGGTATGTCGCGAATACGATAGGATCCGGCTGTTCCGCTGGCGCG 2496
DB 2401 GTTCGGCTGGGGGTATGTCGCGAATACGATAGGATCCGGCTGTTCCGCTGGCGCG 2460
QY 2497 CGATTGTTGGCGCTTTCCGGCCGTTCCCATGCTCTGCGGCGAGCGCAAAATGCCCCGTGA 2556
DB 2461 CGATTGTTGGCGCTTTCCGGCCGTTCCCATGCTCTGCGGCGAGCGCAAAATGCCCCGTGA 2520
QY 2557 ACCGCGTGCAGAAAATCGATTTCTTCGGGTGAAAGTTCGGGGTGGGGCCGGCATCGGCA 2616
DB 2521 ACCGCGTGCAGAAAATCGATTTCTTCGGGTGAAAGTTCGGGGTGGGGCCGGCATCGGCA 2580
QY 2617 CGGGCGATACGAGACGTGCGCTCATGAGTTCTCAAGCGCGCCGCTTATATCGGCGATAG 2676
DB 2581 CGGGCGATACGAGACGTGCGCTCATGAGTTCTCAAGCGCGCGCTTATATCGGCGATAG 2640
QY 2677 GCCTTGGCCATTTCCGGGGCATACATCTCGAAAATCGTCCGCTCGCGGGCGCGATCG 2736
DB 2641 GCCTTGGCCATTTCCGGGGCATACATCTCGAAAATCGTCCGCTCGCGGGCGCGATCG 2700

QY 2737 AACAGCATGCCGACTTCTCTTGGTGTATATCGGGGGGAACTGGAAGCAGGTCTTGAAGCG 2796
DB 2701 AACAGCATGCCGACTTCTCTTGGTGTATATCGGGGGGAACTGGAAGCAGGTCTTGAAGCG 2760
QY 2797 TTGATTTTCGTCTCGGTTCACCGGGCCGTCGATCTTCGCCAGCTTCGCCACAGGGCAACA 2856
DB 2761 TTGATTTTCGTCTCGGTTCACCGGGCCGTCGATCTTCGCCAGCTTCGCCACAGGGCAACA 2820
QY 2857 AGGCGCATGGCGTAAAGCTGATCTCTGTTTCCCGAGGGCCGAGCAATCTTGGCAGCGCG 2916
DB 2821 AGGCGCATGGCGTAAAGCTGATCTCTGTTTCCCGAGGGCCGAGCAATCTTGGCAGCGCG 2880
QY 2917 AAAAAGCGCGCGCTGTGTTGGGATTCGGGATTCGGGGGAAAGCGCTCTACTCCAGCG 2976
DB 2881 AAAAAGCGCGCGCTGTGTTGGGATTCGGGATTCGGGGGAAAGCGCTCTACTCCAGCG 2940
QY 2977 CCCGTTGAGGCTTTAGTGTAGGAAACCGTTATCGCGGATCCCGGCGCTTCGCCCGCATC 3036
DB 2941 CCCGTTGAGGCTTTAGTGTAGGAAACCGTTATCGCGGATTCGCCCGCTTCGCCCGCATC 3000
QY 3037 AGTCTCCGAAGACCAACCAACCGGAGCCCGACACCAACCGGAAACATCTTGCCTCCAG 3096
DB 3001 AGTCTCCGAAGACCAACCAACCGGAGCCCGACACCAACCGGAAACATCTTGCCTCCAG 3060
QY 3097 ATAGCCATGTCTCAACCTTAGCAGCCCGCTCAGAGGGGCAAAATGACAGATCGGAGGCTA 3156
DB 3061 ATAGCCATGTCTCAACCTTAGCAGCCCGCTCAGAGGGGCAAAATGACAGATCGGAGGCTA 3120
QY 3157 GGTGTAGGTGCTGATCGCCAAACCGCGGCTTCGGGTGCTGTAGAGCTAGGAGTTTAC 3216
DB 3121 GGTGTAGGTGCTGATCGCCAAACCGCGGCTTCGGGTGCTGTAGAGCTAGGAGTTTAC 3180
QY 3217 GAACCTATTCGCTCTCATGCTTTTGTAGGGCCAGGTTCTTCTGTTCGTTTCATGACGGAT 3276
DB 3181 GAACCTATTCGCTCTCATGCTTTTGTAGGGCCAGGTTCTTCTGTTCGTTTCATGACGGAT 3240
QY 3277 ATTTTATGCCCCCTTGTATCCAGACTGCTACTTTCGATCCCTTTCGGCTCTGTATGACGAA 3336
DB 3241 ATTTTATGCCCCCTTGTATCCAGACTGCTACTTTCGATCCCTTTCGGCTCTGTATGACGAA 3300
QY 3337 CTGATGATCTTTTGTATCAAGCGTCTCCAATGTGGCTGACAGAAAGTCTCAACTGGTTG 3396
DB 3301 CTGATGATCTTTTGTATCAAGCGTCTCCAATGTGGCTGACAGAAAGTCTCAACTGGTTG 3360
QY 3397 CGGAAGCGGATCAATAATGGTTTCGGATTCGGCGGGCGTGTCTTCATCTCTGGCGGC 3456
DB 3361 CGGAAGCGGATCAATAATGGTTTCGGATTCGGCGGGCGTGTCTTCATCTCTGGCGGC 3420
QY 3457 GTTCTGTCCATCTGCTGCTTCTGCTGTGATGCTGCGCGCTCGGGCTGATGTTGCTT 3516
DB 3421 GTTCTGTCCATCTGCTGCTTCTGCTGTGATGCTGCGCGCTCGGGCTGATGTTGCTT 3480
QY 3517 GCGCAGGATATTCGCTTCTCCGTCGCTTCAGGGCGGCTTCGCTGGCTGATCGAAGT 3576
DB 3481 GCGCAGGATATTCGCTTCTCCGTCGCTTCAGGGCGGCTTCGCTGGCTGATCGAAGT 3540
QY 3577 CAACATCCGATTTGGCTTGGCGCTTCCGCGGAAAAAGCGGAGCAAGCTAACCTTCTGCTGG 3636
DB 3541 CAACATCCGATTTGGCTTGGCGCTTCCGCGGAAAAAGCGGAGCAAGCTAACCTTCTGCTGG 3600
QY 3637 ACGTGTTTCTGAAGATGTGCTAGTGTGCAACCCGAGGGGTGAAGCAGTGGGCGCTCT 3696
DB 3601 ACGTGTTTCTGAAGATGTGCTAGTGTGCAACCCGAGGGGTGAAGCAGTGGGCGCTCT 3660
QY 3697 GGTGCTCGCGGATCGAGAGAGCCACCAAGACAGCAAGAGCTCTGCTGCGGAGCTGCG 3756
DB 3661 GGTGCTCGCGGATCGAGAGAGCCACCAAGAGCTCTGCTGCGGAGCTGCG 3720
QY 3757 GCCATCGCTCCAGTATAGCCCAAGCTTCGGGTTCCAGTTCGACGAGCTCCGCTGTCTCT 3816
DB 3721 GCCATCGCTCCAGTATAGCCCAAGCTTCGGGTTCCAGTTCGACGAGCTCCGCTGTCTCT 3780
QY 3817 GACAGACAGGCTGCGGTTGTGACGAGATCACTCATTCGGGTTGTTTCTCAAGGCGCTTCA 3876

|||||
Db 3781 GACAGACAGCGCTCGTTTCAGGAGATCACTCATTCGGTGTGTTCTCAAGGCGCTTCA 3840
Qy 3877 AAGCCCATGTCGGTTTCGGAACATCAGGTCGGATCACTCAGAGTCCTCCGCGCAG 3936
|||||
Db 3841 AAGCCCATGTCGGTTTCGGAACATCAGGTCGGATCACTCAGAGTCCTCCGCGCAG 3900
Qy 3937 AAGATATAAGCAGCAGGATCGCGCGAGTTGCGGATCGGATCAGGACATGACGACGAC 3996
|||||
Db 3901 AAGATATAAGCAGCAGGATCGCGCGAGTTGCGGATCGGATCAGGACATGACGACGAC 3960
Qy 3997 CGGTTGCGTCCCAATCCGTTTACCGCGAGAGCCAGAAAAAAGCTCCGGAATGTCCGCAATTA 4056
|||||
Db 3961 CGGTTGCGTCCCAATCCGTTTACCGCGAGAGCCAGAAAAAAGCTCCGGAATGTCCGCAATTA 4020
Qy 4057 TCCAGCCGACAGGATCGTCGAGTTTGGTGCATCAGCTCGGCGCGGCGCTG 4109
|||||
Db 4021 TCCAGCCGACAGGATCGTCGAGTTTGGTGCATCAGCTCGGCGCGGCGCTG 4073

RESULT 2
AF027127 AF027127 4643 bp DNA circular SYN 08-NOV-1997
LOCUS Eukaryotic luciferase expression vector pTATALUC+, complete
DEFINITION sequence.
ACCESSION AF027127
VERSION AF027127.1 GI:2598100
KEYWORDS Expression vector pTATALUC+.
SOURCE Expression vector pTATALUC+.
ORGANISM Artificial sequences; vectors.
REFERENCE 1 (bases 1 to 4643)
AUTHORS Altschmid, J. and Duschl, J.
TITLE Set of optimized luciferase reporter gene plasmids compatible with
widely used CMV vectors
JOURNAL BioTechniques 23 (3), 436-438 (1997)
MEDLINE 97443353
PUBMED 9298214
REFERENCE 2 (bases 1 to 4643)
AUTHORS Altschmid, J. and Duschl, J.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-1997) Physiological Chemistry I/BioCenter,
University of Wuerzburg, Am Hubland, Wuerzburg 97074, Germany
FEATURES
Location/Qualifiers
1..4643
/organism="Expression vector pTATALUC+"
/db_xref="taxon:68900"
/plasmid="pTATALUC+"

misc_feature 1..36
/note="polylinker"
misc_feature 37..131
/note="tk-promoter"
gene 170..1822
/gene="luc+"
CDS 170..1822
/gene="luc+"
/codon_start=1
/product="luciferase"
/protein_id="AAB83989.1"
/db_xref="GI:2598101"
/translation="MEDAKNKKGPAPYPYLEDGTAGTQQLHKMKRYALVPGTIAFTD
ARIEVDITYAEFYSFVLAEMKRYGLNTNHRIVVCSSENSLOFFMPVGLGALFIGVAV
APANDYNERELLASMGISQPTVVVFSKKGLOKILNVOKKLPILQKIIIMDSKTDYOG
FSMYFTVTHLPGFENYDFVPESEDRDKTIALIMNSGSGTGLPKGLVPHRTACVR
QSHADPFGNQIIPDAILSVVPFHGFCGFTTGLYLCGFRVVLVYRFEELFLRS
LQDYKIQSALVPLTFSFAKSTLIDKDLNLHEIASGAPLSEKVEGAVAKRPHLP
GRQGYGTETTSATLITPEGDKPKYAVGVVFFFAKVDLDGTGLVNGRGLCV
RPMIMSYVNNPNTALIDKGLHSGDIAYWDEHFFIVDLRLSLIKYKGYQVA
PAELSLILQHPNIFDAGVAGLPDDAGELPAVAVVLEHGKMTKEIKVDVVASQVTT
AKKLGGVVVDEVPKGLTGKLDARKITREILIKAKKGKGIIV"
1854..2075
/note="splice and polyadenylation signals"

misc_feature 2078..2094

/note="polylinker"
complement(3277..4137)
/gene="bla"
complement(3277..4137)
/gene="bla"
/codon_start=1
/product="beta-lactamase"
/protein_id="AAB83990.1"
/db_xref="GI:2598102"
/translation="MSIQHERVALIPFAAFCLPVFAHPETLVKVKDAEDQLGARVGY
IEDLSNGKILESFRPEERPMSTFKVLICGAVLSRIDAGQEQLGRIIRHYSDNLDVE
YSPYTERKLDGTVTRVRELCSAAITMSDNTAANLLTTIGGPKELTAFLNHMGDHTRL
DRWPELNEAIPNDERDTTPVAMATTLRKLITGLLTLASRQQLIDWMEADKVGPL
LRSLAPAGWFTADKSGAGERSGIIAALGPDGKPSRVIVYITGTSQATWMDERNRQIA
EIGASLLIKHW"
misc_feature 4269..4640
/note="polyadenylation signals"
BASE COUNT 1243 a 1078 c 1097 g 1225 t
ORIGIN

Query Match 1.0%; Score 40; DB 12; Length 4643;
Best Local Similarity 100.0%; Pred. No. 9.3e-09;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGCTTGCATCGCTCGAGTCGACTCTAGAGGATCCGGTT 40
|||||
Db 1 AAGCTTGCATCGCTCGAGTCGACTCTAGAGGATCCGGTT 40

RESULT 3
LOCUS SYNIFOKI 91 bp DNA linear SYN 27-APR-1993
DEFINITION MboII/FokI multiple cloning site cassette DNA in plasmid pRBL77.
ACCESSION M30778
VERSION M30778.1 GI:208631
KEYWORDS Synthetic DNA.
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 91)
AUTHORS Hasan, N., Kur, J., and Szybalski, W.
TITLE An MboII/FokI trimming plasmid allowing consecutive cycles of
precise 1- to 12-base-pair deletions in cloned DNA
JOURNAL Gene 82 (2), 305-311 (1989)
MEDLINE 90060819
PUBMED 2583524
FEATURES Location/Qualifiers
source
1..91
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 19 a 26 c 25 g 21 t
ORIGIN

Query Match 0.9%; Score 38; DB 12; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGCTTGCATCGCTCGAGTCGACTCTAGAGGATCCGG 38
|||||
Db 91 AAGCTTGCATCGCTCGAGTCGACTCTAGAGGATCCGG 54

RESULT 4
LOCUS AX256425 360 bp DNA linear PAT 10-OCT-2001
DEFINITION Sequence 9 from Patent WO0076523.
ACCESSION AX256425
VERSION AX256425.1 GI:16075243
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 360)

AUTHORS Forman,B.M., Beard,R.L. and Chandraratna,R.A.
 TITLE Methods for modulating fxr receptor activity
 JOURNAL Patent: WO 0076523-A 9 21-DEC-2000;
 Allergan Sales, Inc. (US)

FEATURES
 source
 1..360
 Location/Qualifiers

BASE COUNT 88 a 103 c 103 g 66 t
 ORIGIN
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Promoter region of plasmid TK-luc"

Query Match 0.9%; Score 38; DB 6; Length 360;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 38
 ||||||||||||||||||||||||||||||||||||||||
 Db 42 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 79

RESULT 5
 DRE238981 416 bp mRNA linear VRT 06-JAN-2000
 LOCUS Danio rerio mRNA for activin beta B protein, partial.
 DEFINITION
 ACCESSION AJ238981
 VERSION AJ238981.1 GI:4867810
 KEYWORDS activin beta B protein.
 SOURCE Danio rerio.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 416)
 AUTHORS Rodaway,A., Takeda,H., Koshida,S., Broadbent,J., Price,B.,
 Smith,J.C., Patient,R. and Holder,N.
 TITLE Induction of the mesoderm in the zebrafish germ ring by yolk
 cell-derived TGF-beta family signals and discrimination of mesoderm
 and endoderm by FGF
 JOURNAL Development 126 (14), 3067-3078 (1999)
 MEDLINE 99307072
 PUBMED 10375499
 REFERENCE 2 (bases 1 to 416)
 AUTHORS Smith,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (14-MAY-1999) Smith J.C., Developmental Biology, National
 Institute for Medical Research, The Ridgeway, Mill Hill, London NW7
 1AA, UNITED KINGDOM

FEATURES
 source
 1..416
 Location/Qualifiers
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 1..416
 /gene="activin beta B"
 <1..>416
 /gene="activin beta B"
 /function="growth factor"
 /codon_start=1
 /product="activin beta B protein"
 /protein_id="CAB43092.1"
 /db_xref="GI:4867811"
 /db_xref="SPTREMBL:Q9W679"
 /translation="GLEQLKACLVQDSRGSGFCEDGNNGGLCCROOFYIDFRILIGNW
 DWIAPGYGNYGCGPCAYMGVPGSASSFTAVVNOYRMGMSPGSVNSCCIPTK
 LSTWSMLVFDDENVILKRDVPMVEGCGEFFDI"
 BASE COUNT 97 a 107 c 113 g 99 t
 ORIGIN

Query Match 0.9%; Score 38; DB 5; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 38

Db 16 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 53
 ||||||||||||||||||||||||||||||||||||||||

RESULT 6.

LOCUS 109132

DEFINITION Sequence 1 from Patent WO 8900193.

ACCESSION 109132

VERSION 109132.1

KEYWORDS 1 GI:588159

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1195)

AUTHORS Stalker,D.

TITLE HALOARYLNITRILE DEGRADING GENE, ITS USE, AND CELLS CONTAINING THE

JOURNAL GENE

FEATURES Patent: WO 8900193-A 1 12-JAN-1989;

Location/Qualifiers

1..1195

source /organism="unknown"

BASE COUNT 292 a 335 c 330 g 238 t

ORIGIN

Query Match 0.9%; Score 38; DB 6; Length 1195;

Best Local Similarity 100.0%; Pred. No. 1.1e-07;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 38

||||||||||||||||||||||||||||||||||||||

Db 18 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 55

RESULT 7

LOCUS MTGLMS

DEFINITION Mycobacterium tuberculosis glms gene.

ACCESSION AJ000333

VERSION AJ000333.1

KEYWORDS glms gene; glucosamine fructose-6-phosphate aminotransferase.

SOURCE Mycobacterium tuberculosis.

ORGANISM Mycobacterium tuberculosis.

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Corynebacterineae; Mycobacteriaceae;

Mycobacterium tuberculosis complex.

REFERENCE 1 (bases 1 to 2442)

AUTHORS McKendree,W.L.

TITLE Direct Submission

JOURNAL Submitted (08-JUL-1997) McKendree W.L., Subtropical Insects

Research Management Unit, United States Horticultural Research

Laboratory, 2120 Camden Road, Florida, 32803, USA

REFERENCE 2 (bases 1 to 2442)

AUTHORS McKendree,W.L., Schuster,S.M. and Richards,N.

TITLE Structure of a recombinant glms protein from Mycobacterium

tuberculosis expressed in E. coli

JOURNAL Unpublished

FEATURES Location/Qualifiers

1..2442

source /organism="Mycobacterium tuberculosis"

/db_xref="taxon:1773"

/clone="genomic"

/clone_lib="cosmid"

1..2442

gene /gene="glms"

1..2442

/gene="glms"

/function="glucosamine fructose-6-phosphate

aminotransferase"

/codon_start=1

/transl_table=11

/protein_id="CAA04007.1"

/db_xref="GI:2275059"

```

/db_xref="sptrembl:033274"
/translation="MTMITPSLHACRSTLEDPGHPSTLAVTWRMGPGVPRRTVPVTOG
RPGRVWIPARRYSGPGWMDGVRVNGRPGVGGSHGPPNRVIEVGNORDPVRHREERH
OTPTNEQVCLRELVRFEALLARPBGHVGALLPODORTVRPODELQVLGLWGVOLA
FRARWPNSTIASVGRRCRCPDHAALPWSAVCGIVGYVRRPAIVYVMDALRRMEYRGY
DSSGIALVDGTLVRRRAGRLANLEEAEMPTALSGTITGLGHTRWATHGRPTDRN
ADHRDAGGIAVYVHNGIIEENFAVRLRELETAGVEFASTDTEVAHHLVARVYRHGET
ADPFGSVLAVLRLEGHFTLVFANADDPGTLVAAARSTPLVLGIGDNEMFVGSVDAA
FIEHTRAVELGQDAVVITADGVYISDFDNDGLQAGRDFRPHIDWDLAARAEKGGY
EYFMLEKTAQPAQAAVITLGHGVGRIVLDEORLSDOELREIDKVFVACGTAYHSG
LLKALAEHTRLPVEVELASEFRYROPVLDSTLVVAISQSGETADTLEAVRHAKQ
KARYLAICNTNGSIQPRECDALVYTRAGPEIGVASTKTFEAQIAANTLLGLALAQAAG
TKYDVEYREYHELEAMPDLVARVIAATGPVLAELAHFAQSTVFLGRHVGPVVALE
GALKKLKELAYMHAEGFAAGELKHGPIALIEDGLPVIWMPSPKGSATLHAKLSNIRE
IOTRGAVTIVTAEGDETVPYADHLIEIPAVSTLLQPLLSTIPLQVFAASVARARGY
DVDKPRNLAKSVTVE"
BASE COUNT 415 a 800 c 815 g 412 t
ORIGIN
Query Match 0.9%; Score 38; DB 1; Length 2442;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCGG 38
|||||
Db 18 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCGG 55

RESULT 8
SYNCTATBLB 4404 bp DNA linear SYN 16-FEB-1994
LOCUS
DEFINITION
Chloramphenicol acetyltransferase (CAT) and beta-lactamase genes,
complete cds.
ACCESSION M80483
VERSION M80483.1 GI:208101
KEYWORDS
CAT vector; beta-lactamase; chloramphenicol acetyltransferase;
thymidine kinase promoter; transient transfection vector.
SOURCE
Cloning vector DNA.
ORGANISM
artificially cloning vector
unidentified sequences; vectors.
REFERENCE
1 (bases 1 to 4404)
AUTHORS Boshart,M., Kluppel,M., Schmidt,A., Schutz,G. and Luckow,B.
TITLE Reporter constructs with low background activity utilizing the cat
gene
JOURNAL Gene 110 (1), 129-130 (1992)
MEDLINE 92184107
PUBMED 1544570
FEATURES
source
Location/Qualifiers
1..4404
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
31..200
242..901
/organism="CAT"
242..901
/organism="CAT"
/codon_start=1
/transl_table=11
/product="chloramphenicol acetyltransferase"
/protein_id="AA073138.1"
/db_xref="GI:208102"
/translation="MEKKITGYTTVDISQMRHKEHFEAFQSVQACTYNOTVOLDITAF
SLTKVKKHKEYPAFIHLARLMAHPEFRMAMKDGELVINDSVHPCTVTFHEQTETP
LKLKSEYHDDFROFLHIYSDVACYGENLAYEPKGFENMFVSANPWVSTFDLAV
AMNDFPAPVTMGKYYTQGDVKVLMPLAIQVHVAACDGFHVGRMLNELQYCDDEWQGG
A"
991..1604
/misc_feature
/note="SV40 small T splice site"
1601..1841
/polyA_signal
/note="SV40 polyadenylation signal"
complement(3038..3898)
/gene="bla"

```

```

complement(3038..3898)
/gene="bla"
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="AA073139.1"
/db_xref="GI:453366"
/translation="MSIOHFRVALIPFFAFCPLPFAHPETLVKVKDAEDOLGARVG
YEDLSNGHKLEFRPEERPMSTFKVLICGAVLSRIDAGQQLGRRIRHNSQDLVE
YSVTEKHLTDGMTVRELCSAAJTMSDNTAANLLLTIGGPKELTAFLNHMGHVTPL
DRPEPELNEAIPNDRDRTTPVAMATTLRLLTGELLTLASRQOLIDMEADKVGPL
LRGALLPAGFIADKSGAGERSGRIIALLGPDGKPSRIVVIYTTGSAOTWDERNRQIA
EIGASLIKHW"
polyA_signal 4030..4164
/note="SV40 polyadenylation signal"
polyA_signal 4161..4401
/note="SV40 polyadenylation signal"
BASE COUNT 1196 a 1005 c 976 g 1227 t
ORIGIN
Query Match 0.9%; Score 38; DB 12; Length 4404;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCGG 38
|||||
Db 1 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCGG 38

RESULT 9
BLCAT2DNA 4496 bp DNA linear SYN 26-FEB-1992
LOCUS
DEFINITION
Plasmid pBLCAT2 gene for beta-lactamase and CAT gene for
chloramphenicol acetyltransferase.
X64410
X64410.1 GI:58160
KEYWORDS
beta-lactamase; cat gene; chloramphenicol acetyltransferase;
thymidine kinase promoter.
SOURCE
synthetic construct.
artificial sequences.
1 (bases 1 to 4496)
REFERENCE
Luckow,B.H.R.
AUTHORS Direct Submission
TITLE Submitted (06-FEB-1992) B.H.R. Luckow, German Cancer Res Center, Im
Neuenheimer Feld 280, W-6900 Heidelberg, FRG
JOURNAL
REFERENCE
2 (bases 1 to 4496)
AUTHORS Luckow,B. and Schutz,G.
TITLE CAT constructions with multiple unique restriction sites for the
functional analysis of eukaryotic promoters and regulatory elements
Nucleic Acids Res. 15 (13), 5490 (1987)
JOURNAL 87260024
MEDLINE 3037497
PUBMED
COMMENT CAT vector for transient transfection into eukaryotic cells.
Contains tk promoter for analysis of cis-acting elements in front
of a heterologous promoter.
FEATURES
source
Location/Qualifiers
1..4496
/organism="synthetic construct"
/db_xref="taxon:32630"
429..598
/promoter
/note="Herpes simplex virus tk promoter"
640..1299
/gene="CAT"
640..1299
/gene="CAT"
/codon_start=1
/transl_table=11
/product="chloramphenicol acetyltransferase"
/protein_id="CAA45754.1"
/db_xref="GI:58161"
/translation="MEKKITGYTTVDISQMRHKEHFEAFQSVQACTYNOTVOLDITAF
LTKVKKHKEYPAFIHLARLMAHPEFRMAMKDGELVINDSVHPCTVTFHEQTETP

```



```

SSLMSEYHDDFRQFLHIYSODVACYGENLAYFPKGF1ENMFVFSANPWVSFTSLNVA
ANMNOFFAPVETMGKYYTQGDVKVLMPLAIQVHHAVCDFHVGRLMNELOQYCDWEQGG
A"
1389..2002
/Note="SV40 small T intron"
/Note="SV40 signals"
/Note="SV40 signals"
complement(3436..4296)
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="CAA45755.1"
/db_xref="GI:58162"
/translation="MSIQHFRVALIPFAAFCPLPVFAHPETLVKVKDAEDOLGARVGY
IELDLSNGKILSEFRPEERFPMSTFKVLLCGAVLSRIDAGEOLGRIHYSONDLVE
YSPVTEKHLTDGTMVRELCSAAITMSDNTAANLLLTITIGGPKELTAFLLHNGMDHVTRL
DRWPELNEALPNDERTMPVAMATIRLLKLTIGELLTLLASRQQLIDWMEADKVGAPL
LRSALPAGWFIADKSGAGERSGRLIAALGPDGPKPSRIIVYITTGSOATMDERNRQIA
EIGASLIKHW"
BASE COUNT 1188 a 1045 c 1070 g 1193 t
ORIGIN
Query Match 0.9%; Score 38; DB 12; Length 4496;
Best Local Similarity 100.0%; Pred. No. le-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCGG 38
|||||
Db 399 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCGG 436

RESULT 10
AF027128 4711 bp DNA circular SYN 08-NOV-1997
LOCUS Eukaryotic luciferase expression vector ptkLUC+, complete sequence.
DEFINITION AF027128
ACCESSION AF027128
VERSION AF027128.1 GI:2598103
KEYWORDS Expression vector ptkLUC+.
SOURCE Expression vector ptkLUC+
ORGANISM artificial sequences; vectors.
REFERENCE 1 (bases 1 to 4711)
AUTHORS Altschmid,J. and Duschl,J.
TITLE Set of optimized luciferase reporter gene plasmids compatible with
widely used CAT vectors
JOURNAL Biotechniques 23 (3), 436-438 (1997)
MEDLINE 97443353
PUBMED 9298214
REFERENCE 2 (bases 1 to 4711)
AUTHORS Altschmid,J. and Duschl,J.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-1997) Physiological Chemistry I/Biocenter,
University of Wuerzburg, Am Hubland, Wuerzburg 97074, Germany
FEATURES
source
1..4711
/organism="Expression vector ptkLUC+"
/db_xref="taxon:68901"
/plasmid="ptkLUC+"
1..36
/Note="polylinker"
37..200
/Note="tk-promoter"
238..1890
/gene="luc+"
238..1890
/gene="luc+"
/product="luc+"
/codon_start=1
/product="luciferase"
/protein_id="AAB83991.1"
/db_xref="GI:2598104"
/translation="MEDAKNIKGPAPFPYPLEDGTAGEQLHAKMKRYALVPCTIAFTD
AHTEVDITYAEYFEMSVLAEMARYGLNTNHRIVVCSNSLQFFNPVLGALFIGVAV
APANDIYNERELLNSMGISQPTVVVFSKKGLOKILNVQKILPITIKIIMDSKTDYOG

```

```

FQSMYTFVTSHPGFEYDEVPESFORDKTIALIMSSGGTGLPKGVALPHRTACVR
FSHARDPIFGNOIIPDTAILSVVPHHGFMTTGLYLICGFRVVLVYRFEELFLRS
LQDYKIOSALLVPTLFSFAKSTLDKYDLNLHEIASGGAPLSKEVGEANAKRFLP
GIRGYGLTETSAILITPEGDDAPCAVKVVPFEAKVVDLDTGKTLCVQNGELCV
RGMIMSGYVNNPATNALIDKDWLHSGDIAYWDEDEHFFIVDLKLSLIYKGYQVA
PAEELSILLOHPNIFDAGVAGLPDDAGELFAAVVVLHGHKTMTEKEIVDVVASOVTT
AKLRGGVVFVDEVPKGLTGKLDARKIREILIKAKKGGKIAV"
1922..2143
/Note="splice and polyadenylation signals"
2146..2162
/Note="polylinker"
complement(3345..4205)
/gene="bla"
complement(3345..4205)
/gene="bla"
/codon_start=1
/product="beta-lactamase"
/protein_id="AAB83992.1"
/db_xref="GI:2598105"
/translation="MSIQHFRVALIPFAAFCPLPVFAHPETLVKVKDAEDOLGARVGY
IELDLSNGKILSEFRPEERFPMSTFKVLLCGAVLSRIDAGEOLGRIHYSONDLVE
YSPVTEKHLTDGTMVRELCSAAITMSDNTAANLLLTITIGGPKELTAFLLHNGMDHVTRL
DRWPELNEALPNDERTMPVAMATIRLLKLTIGELLTLLASRQQLIDWMEADKVGAPL
LRSALPAGWFIADKSGAGERSGRLIAALGPDGPKPSRIIVYITTGSOATMDERNRQIA
EIGASLIKHW"
4337..4708
/Note="polyadenylation signals"
BASE COUNT 1253 a 1102 c 1120 g 1236 t
ORIGIN
Query Match 0.9%; Score 38; DB 12; Length 4711;
Best Local Similarity 100.0%; Pred. No. le-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCGG 38
|||||
Db 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCGG 38

RESULT 11
AF102233/3 AF102233/3 7055 bp. DNA circular SYN 24-NOV-1998
LOCUS Transposon delivery vector putkml, complete sequence.
DEFINITION AF102233
ACCESSION AF102233
VERSION AF102233.1 GI:3907623
KEYWORDS Transposon delivery vector putkml.
SOURCE Transposon delivery vector putkml.
ORGANISM artificial sequences; vectors.
REFERENCE 1 (bases 1 to 7055)
AUTHORS de Lorenzo,V., Herrero,M., Jakubzik,U. and Timmis,K.N.
TITLE Mini-trn5 transposon derivatives for insertion mutagenesis, promoter
probing, and chromosomal insertion of cloned DNA in gram-negative
eubacteria
JOURNAL J. Bacteriol. 172 (11), 6568-6572 (1990)
MEDLINE 91035272
PUBMED 2172217
REFERENCE 2 (bases 1 to 7055)
AUTHORS Tombolini,R. and Jansson,J.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1998) Department of Biochemistry, Stockholm
University, Svante Arrhenius vag. 10-12, Stockholm SE-10691, Sweden
FEATURES
Location/Qualifiers
1..7055
/organism="Transposon delivery vector putkml"
/db_xref="taxon:84477"
complement(311..1171)
/gene="bla"
complement(311..1171)
/gene="bla"
/codon_start=1
/transl_table=11
/product="beta lactamase"

```



```

BASE COUNT      2331 a 1764 c 1765 g 2282 t
ORIGIN
Query Match      0.9%; Score 38; DB 12; Length 8142;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 38
|||||
Db 133 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 170

RESULT 13
XXU02428/c      10737 bp      DNA      circular SYN 29-JAN-1997
LOCUS
DEFINITION      Cloning vector pDR2, complete sequence.
ACCESSION      U02428
VERSION      U02428.1 GI:413794
KEYWORDS
SOURCE
ORGANISM
Cloning vector pDR2.
artificial sequences: vectors.
REFERENCE
AUTHORS      Murphy,A.J., Kung,A.L., Swirski,R.A. and Schinke,R.T.
TITLE      cDNA expression cloning in human cells using the plambdaDR2
JOURNAL
METHODS      A Companion to Methods in Enzymology 4, 111-131 (1992)
FEATURES
source
Location/Qualifiers
/organism="Cloning vector pDR2"
/db_xref="taxon:31805"
/notice="plasmid released from lambda DR2"
BASE COUNT      2723 a 2615 c 3053 g 2459 t
ORIGIN
Query Match      0.9%; Score 38; DB 12; Length 10850;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 38
|||||
Db 782 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 745

RESULT 15
A25917/c
LOCUS
DEFINITION      Expression vector sequence.
ACCESSION      A25917
VERSION      A25917.1 GI:833587
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
FEATURES
source
Location/Qualifiers
1..92
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT      25 a 20 c 25 g 22 t
ORIGIN
Query Match      0.9%; Score 37; DB 6; Length 92;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCG 37
|||||
Db 47 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCG 11

Search completed: May 19, 2003, 07:44:01
Job time : 10419 secs

```

```

BASE COUNT      2331 a 1764 c 1765 g 2282 t
ORIGIN
Query Match      0.9%; Score 38; DB 12; Length 8142;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 38
|||||
Db 133 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 170

RESULT 13
XXU02428/c      10737 bp      DNA      circular SYN 29-JAN-1997
LOCUS
DEFINITION      Cloning vector pDR2, complete sequence.
ACCESSION      U02428
VERSION      U02428.1 GI:413794
KEYWORDS
SOURCE
ORGANISM
Cloning vector pDR2.
artificial sequences: vectors.
REFERENCE
AUTHORS      Murphy,A.J., Kung,A.L., Swirski,R.A. and Schinke,R.T.
TITLE      cDNA expression cloning in human cells using the plambdaDR2
JOURNAL
METHODS      A Companion to Methods in Enzymology 4, 111-131 (1992)
FEATURES
source
Location/Qualifiers
/organism="Cloning vector pDR2"
/db_xref="taxon:31805"
/notice="plasmid released from lambda DR2"
BASE COUNT      2723 a 2615 c 3053 g 2459 t
ORIGIN
Query Match      0.9%; Score 38; DB 12; Length 10850;
Best Local Similarity 100.0%; Pred. No. 1e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 38
|||||
Db 782 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCGG 745

RESULT 15
A25917/c
LOCUS
DEFINITION      Expression vector sequence.
ACCESSION      A25917
VERSION      A25917.1 GI:833587
KEYWORDS
SOURCE
ORGANISM
synthetic construct.
artificial sequences.
FEATURES
source
Location/Qualifiers
1..92
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT      25 a 20 c 25 g 22 t
ORIGIN
Query Match      0.9%; Score 37; DB 6; Length 92;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCG 37
|||||
Db 47 AAGCTTGATCCCTGCAGGTGCTAGAGGATCCG 11

Search completed: May 19, 2003, 07:44:01
Job time : 10419 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: May 19, 2003, 01:52:01 ; Search time 844 Seconds
(without alignments)
10979.826 Million cell updates/sec

Title: US-09-926-163B-1

Perfect score: 4115

Sequence: 1 aagcttgatcgtcgcaggt.....tcggggcgccgtgaagctt 4115

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 2185239 seqs, 1125999159 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4115	100.0	4115	21 AAA97430	Gluconobacter oxyd
2	38	0.9	1195	10 AAN91189	Modified nitrilas
3	37	0.9	192	12 AAQ11697	Plasmid pWH332 fra
4	37	0.9	116	21 AAA10800	Nucleotide sequenc
5	37	0.9	586	12 AAQ11823	Region of pUC19-SA
6	37	0.9	586	15 AAQ78008	Bacteriophage T3 S
7	37	0.9	586	16 AAQ87073	Bacteriophage T3 S
8	37	0.9	586	17 AAT09985	S-adenosylmethioni
9	37	0.9	586	17 AAT09984	S-adenosylmethioni

10	37	0.9	586	18 AAT58564	DNA sequence of Ad
11	37	0.9	586	20 AAA00468	S-adenosylmethioni
12	37	0.9	3796	21 AAA27831	Vector plasmid pCM
13	37	0.9	4910	17 AAT15287	prEP78/52 hybrid a
14	37	0.9	6295	19 AAV02043	Plasmid pWRG3196 e
15	37	0.9	7164	22 AAD10238	Commercial plasmid
16	37	0.9	7380	20 AA840428	MMP9 promoter beta
17	37	0.9	7383	22 AAS00153	Matrix metalloprot
18	37	0.9	7612	21 AA396629	DNA sequence of pl
19	37	0.9	7897	17 AAT27555	Shuttle vector pAD
20	37	0.9	8540	16 AAT08476	Eukaryotic express
21	37	0.9	8540	17 AAT41900	Vector pAPEX-3p.
22	37	0.9	8932	20 AAA19816	Apex-3 eukaryotic
23	37	0.9	8932	21 AAA07342	Vector pAPEX-3. S
24	37	0.9	9972	17 AAT27557	Shuttle vector pAD
25	37	0.9	19307	17 AAT27558	Shuttle vector pAD
26	36	0.9	40	18 AAT46199	Capture DNA oligom
27	36	0.9	50	16 AAQ96167	Stabiliser sequenc
28	36	0.9	50	24 ABU51437	Bombyx mori R2 ele
29	36	0.9	57	13 AAQ25912	pUC18 polylinker (
30	36	0.9	59	21 AAA50624	Phage M13 template
31	36	0.9	59	22 AAF82286	Lux plasmid pGLSL1
32	36	0.9	60	21 AAA90483	Plasmid pUC18 mult
33	36	0.9	60	22 AAF74634	Cloning vector pHK
34	36	0.9	60	24 ABA99032	Plasmid pECO2.0 mu
35	36	0.9	63	21 AAC63087	Mutant cre pools p
36	36	0.9	66	15 AAQ66558	Streptavidin/prote
37	36	0.9	66	24 ABU59115	Nucleotide sequenc
38	36	0.9	71	19 AAV71993	Cloned VNTR allele
39	36	0.9	72	17 AAT17251	Recombinase enzyme
40	36	0.9	72	17 AAT17252	Recombinase enzyme
41	36	0.9	72	18 AAT65088	SacI-HindIII regio
42	36	0.9	72	21 AA399200	Multicloning site
43	36	0.9	74	14 AAQ37537	Potato virus X alp
44	36	0.9	74	15 AAQ64588	M13mpl8(+) univers
45	36	0.9	75	24 ABR66942	Synthetic plasmid

ALIGNMENTS

RESULT 1
AAA97430
ID AAA97430 standard; DNA; 4115 BP.

-XX AC AAA97430;

XX XX

DT 29-JAN-2001 (first entry)

XX XX Gluconobacter oxydans D-sorbitol dehydrogenase (SLDH) gene.

DE DE

XX XX D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;

KW KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;

KW KW L-ascorbic acid biosynthesis; vitamin C; ds.

XX XX Gluconobacter oxydans.

XX XX WO200055329-A1.

XX XX 21-SEP-2000.

XX XX 16-MAR-2000; 2000WO-JP01608.

XX XX 17-MAR-1999; 99JP-0072810.

XX XX 06-AUG-1999; 99JP-0224679.

XX XX (FUJI) FUJISAWA PHARM CO LTD.

XX XX Shibata T., Ichikawa C, Matsuura M, Noguchi Y, Saito Y;

PI PI Yamashita M, Takata Y;

XX XX WPI; 2000-587530/55.

DR DR P-PSDB; AAB23172.

Db 1681 AGGTGTTCTGGACGGAACCGCTGCGCTCGGGCGATCGAAGACGCGGACCTGTCAACGTA 1740
QY 1741 TAGCCTTCGGAAATTCATCTATCTCGAAATGCTCGTGGTGGCGACAGAGGCGGGGA 1800
Db 1741 TAGCCTTCGGAAATTCATCTATCTCGAAATGCTCGTGGTGGCGACAGAGGCGGGGA 1800
QY 1801 CGTATGAATCGTCCGAGCGGACTTATGCGGACGCGGGAATGGAAGTTGGCCAAAGCGGACG 1860
Db 1801 CGTATGAATCGTCCGAGCGGACTTATGCGGACGCGGGAATGGAAGTTGGCCAAAGCGGACG 1860
QY 1861 ACTTCGAAGGCTCTCTGAAGCTCCCGGCTTCGATGGGTGGCGGATCTGGATACGTCGG 1920
Db 1861 ACTTCGAAGGCTCTCTGAAGCTCCCGGCTTCGATGGGTGGCGGATCTGGATACGTCGG 1920
QY 1921 AACTGGATCAAAAGGTATCTGCTGCGGAAGATCATCCGGGAAAGGCGGTAAAGCGG 1980
Db 1921 AACTGGATCAAAAGGTATCTGCTGCGGAAGATCATCCGGGAAAGGCGGTAAAGCGG 1980
QY 1981 CCATCCCGGCTGAATTCGGCTTTTAGGGTAGCGACTGAAACAGAAAACCGCGCTCTGGA 2040
Db 1981 CCATCCCGGCTGAATTCGGCTTTTAGGGTAGCGACTGAAACAGAAAACCGCGCTCTGGA 2040
QY 2041 AGGACGCGGTTTTTTTTATGCTCAGATCTGTCCATCAGGACAGGATCAGGACCA 2100
Db 2041 AGGACGCGGTTTTTTTTATGCTCAGATCTGTCCATCAGGACAGGATCAGGACCA 2100
QY 2101 CGATCAGGACAAAGTCCGCTGGAGGGGAGCCCCATTTCGAACTGTACGCCCATGACGGCA 2160
Db 2101 CGATCAGGACAAAGTCCGCTGGAGGGGAGCCCCATTTCGAACTGTACGCCCATGACGGCA 2160
QY 2161 GCGACCGAGATCAGGATTACAAGAGGATCAGTCCCATGGACATCTCTTTGCGCGGTT 2220
Db 2161 GCGACCGAGATCAGGATTACAAGAGGATCAGTCCCATGGACATCTCTTTGCGCGGTT 2220
QY 2221 GAGACTGCTGTGTTCGGGTGTCTAAAAGTTTCGTAGGGGCGGAAAGATCAAGC 2280
Db 2221 GAGACTGCTGTGTTCGGGTGTCTAAAAGTTTCGTAGGGGCGGAAAGATCAAGC 2280
QY 2281 TGTGGTCCGCTTAATCCGTTCCCAAGCCGATTTGATGCGGGCCACCCGCTCTGTGCG 2340
Db 2281 TGTGGTCCGCTTAATCCGTTCCCAAGCCGATTTGATGCGGGCCACCCGCTCTGTGCG 2340
QY 2341 CGTTTGGCCTCTGCTCTGACATAGTTTCTGGGCCAGCAGCTCCGATGATGTTCCGGG 2400
Db 2341 CGTTTGGCCTCTGCTCTGACATAGTTTCTGGGCCAGCAGCTCCGATGATGTTCCGGG 2400
QY 2401 ATCAGGTTGCGCCAGCGCACCGGATTTCTGTGTCAGTTGCGCTGCGGGTGATGCCGAGA 2460
Db 2401 ATCAGGTTGCGCCAGCGCACCGGATTTCTGTGTCAGTTGCGCTGCGGGTGATGCCGAGA 2460
QY 2461 ATAGATAGGATCCGGCTCGTTTCCGCTGGCGGCGGATTTGTTGCCGTTTCCGGCCCGG 2520
Db 2461 ATAGATAGGATCCGGCTCGTTTCCGCTGGCGGCGGATTTGTTGCCGTTTCCGGCCCGG 2520
QY 2521 TCCCATGCTCTGCGGCGAGCCCAATGCCCCGCTGAGCGCTCGAGAAAATCGATTCC 2580
Db 2521 TCCCATGCTCTGCGGCGAGCCCAATGCCCCGCTGAGCGCTCGAGAAAATCGATTCC 2580
QY 2581 TTCGGGTGAAGCTCGCGGCTGGGCGGCATCTGGCACGGCGGATACGGAACAGTCCCGTC 2640
Db 2581 TTCGGGTGAAGCTCGCGGCTGGGCGGCATCTGGCACGGCGGATACGGAACAGTCCCGTC 2640
QY 2641 ATGAGGTTCTCAAGCGGCGCGTATTATCGGCATAGGCTTGGCCATTTCGCGGGCATAC 2700
Db 2641 ATGAGGTTCTCAAGCGGCGCGTATTATCGGCATAGGCTTGGCCATTTCGCGGGCATAC 2700
QY 2701 ATCTCGAATCGTCCGCTCGGGCGCGGATCGAAGACAGATCCGACTTCCTTGGTG 2760
Db 2701 ATCTCGAATCGTCCGCTCGGGCGCGGATCGAAGACAGATCCGACTTCCTTGGTG 2760
QY 2761 TTATCGGGGGGAACGTGAAGCAGGCTTTGAAAGCGTTGATTCTGTCGCTTACCCGGC 2820
Db 2761 TTATCGGGGGGAACGTGAAGCAGGCTTTGAAAGCGTTGATTCTGTCGCTTACCCGGC 2820

QY 2821 CCGTCGATCTTTCGCCAGCTTCGCGCACAGGGCAACAGGCCGATGGCGTAAAGCTGATCT 2880
Db 2821 CCGTCGATCTTTCGCCAGCTTCGCGCACAGGGCAACAGGCCGATGGCGTAAAGCTGATCT 2880
QY 2881 CGTTTGGCCAGGGCGGAGCAATCTTGGCAGCGCGGAAAAAGGCGCGCTGTTGGGATCG 2940
Db 2881 CGTTTGGCCAGGGCGGAGCAATCTTGGCAGCGCGGAAAAAGGCGCGCTGTTGGGATCG 2940
QY 2941 GGAGCGCATTTCCGGGAAAGCGCTCACTCCAGCGCGCGTTGAGGGCTTCAGTAGCCAA 3000
Db 2941 GGAGCGCATTTCCGGGAAAGCGCTCACTCCAGCGCGCGTTGAGGGCTTCAGTAGCCAA 3000
QY 3001 CCGTTATCGGCGGATGCCCGAGCGCTCGGCCCATCATAGTCTCCGAAAAAGGACCACCAACC 3060
Db 3001 CCGTTATCGGCGGATGCCCGAGCGCTCGGCCCATCATAGTCTCCGAAAAAGGACCACCAACC 3060
QY 3061 GCGAAGCCCGGACACCAACCGCAATCTTGGCCCGAGATAGGCATGTCTCAACCTAGCAC 3120
Db 3061 GCGAAGCCCGGACACCAACCGCAATCTTGGCCCGAGATAGGCATGTCTCAACCTAGCAC 3120
QY 3121 GCCCGCTCACAGCGGCAATCAGATCGCAGGCTAGGTGATAGGTGCTGATGCCCAACC 3180
Db 3121 GCCCGCTCACAGCGGCAATCAGATCGCAGGCTAGGTGATAGGTGCTGATGCCCAACC 3180
QY 3181 GCCCGGCTTCGGGTGTGTAGAGCTAGGAGTTACGAACCTTATCGCTGTCTCATGCTTT 3240
Db 3181 GCCCGGCTTCGGGTGTGTAGAGCTAGGAGTTACGAACCTTATCGCTGTCTCATGCTTT 3240
QY 3241 TGAGCGCAGGTTCTTCTGTTCTCATGACGGATATTTTATGCCACCTTGATCCAG 3300
Db 3241 TGAGCGCAGGTTCTTCTGTTCTCATGACGGATATTTTATGCCACCTTGATCCAG 3300
QY 3301 ACTGCTACTTCGATCCCTTCCGCTCTGATCAGCAACTGATGGATCTTTTATCAAGCGT 3360
Db 3301 ACTGCTACTTCGATCCCTTCCGCTCTGATCAGCAACTGATGGATCTTTTATCAAGCGT 3360
QY 3361 CTGCAATGTGGCTGCAGAAAGTGTGAACCTGTTGCGGGAAGCGGATCAATAATGGGTT 3420
Db 3361 CTGCAATGTGGCTGCAGAAAGTGTGAACCTGTTGCGGGAAGCGGATCAATAATGGGTT 3420
QY 3421 CGGATTCGGGCGGGGCTGTCTTTCATGCTGGCGGCGCTTCTGTCATCTCCCTGTTCTG 3480
Db 3421 CGGATTCGGGCGGGGCTGTCTTTCATGCTGGCGGCGCTTCTGTCATCTCCCTGTTCTG 3480
QY 3481 GGTCTGTGATGCTGCGGCTCGGCTGATGTTGCTGCGCAGGATATTCCTTCTCCGT 3540
Db 3481 GGTCTGTGATGCTGCGGCTCGGCTGATGTTGCTGCGCAGGATATTCCTTCTCCGT 3540
QY 3541 GCGCTTCAGGCGCGCTTTCGCTGCTGGATCGAACGTCACATCCGGAATGGCTGGCGCTT 3600
Db 3541 GCGCTTCAGGCGCGCTTTCGCTGCTGGATCGAACGTCACATCCGGAATGGCTGGCGCTT 3600
QY 3601 CCGCGGAAAAAGCGGCAAGCTAACCCCTTCGCTGGAGCTGTTTCTGAAGATGTCTCAGT 3660
Db 3601 CCGCGGAAAAAGCGGCAAGCTAACCCCTTCGCTGGAGCTGTTTCTGAAGATGTCTCAGT 3660
QY 3661 GCTCAACCCCGAGGGCTGAAGCCAGTGGGCGCTCTGGTGGTTCGCGGCGATCCAGAGAA 3720
Db 3661 GCTCAACCCCGAGGGCTGAAGCCAGTGGGCGCTCTGGTGGTTCGCGGCGATCCAGAGAA 3720
QY 3721 GCCACCGAGACGCAAAAGCTTCTGCTGGGCGACTCGGCCATCGGCTCCAGTATAGCCCA 3780
Db 3721 GCCACCGAGACGCAAAAGCTTCTGCTGGGCGACTCGGCCATCGGCTCCAGTATAGCCCA 3780
QY 3781 AACTCGGGTTCAGTGCACGAGCTCGGCTGCTGACAGACAGCGCTCGGTTTGACG 3840
Db 3781 AACTCGGGTTCAGTGCACGAGCTCGGCTGCTGACAGACAGCGCTCGGTTTGACG 3840
QY 3841 AGATCACTCATTCGGGTTGTTTCTCAAGGCGCTTCAAGGCCCATTTGCGGGTTTCGGAAA 3900
Db 3841 AGATCACTCATTCGGGTTGTTTCTCAAGGCGCTTCAAGGCCCATTTGCGGGTTTCGGAAA 3900

XX PD 26-JAN-2000.
 XX XX
 XX PF 17-JUL-1998; 98CN-0116333.
 XX DR
 XX PR 17-JUL-1998; 98CN-0116333.
 XX PA (HUAC-) HUACHEN BIOLOGICAL TECH RES INST SHANGHA.
 XX XX
 XX PI Cao X, Zhao Z, Huang X;
 XX XX WPI: 2000-351200/31.
 XX DR
 XX XX Generation, cloning and sequence determination of a vector, and the
 XX PT use thereof
 XX XX
 XX PS Example 1; Fig 3; 16pp; Chinese.
 XX XX
 XX CC This sequence represents the nucleotide sequence of pJL6. The invention
 XX CC relates to methods for producing a vector containing the following
 XX CC elements:
 XX CC (i) a promoter for controlling exogenous gene transcription;
 XX CC (ii) a reverse sequencing primer sequence;
 XX CC (iii) an RBS (ribosome binding site) sequence;
 XX CC (iv) a multiple cloning site directly downstream of the SD sequence;
 XX CC (v) a positive sequencing primer sequence;
 XX CC (vi) a terminator for controlling exogenous gene termination.
 XX CC The multiple cloning site contains the enzymic incision site of a blunt
 XX CC end type restriction endonuclease, and the interval of SD-AUG is 5-8
 XX CC nucleotides. This method may be used to produce other vectors for
 XX CC sequencing, cloning, and the expression of exogenous genes in the
 XX CC molecular biology field.
 XX XX
 XX SQ Sequence 116 BP; 32 A; 24 C; 28 G; 32 T; 0 other;
 Query Match 0.9%; Score 37; DB 21; Length 116;
 Best Local Similarity 100.0%; Pred. No. 4.2e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AAGCTTCATGCGCTGCAGTGCAGTCTAGAGGATCCG 37
 |||||
 DB 91 AAGCTTCATGCGCTGCAGTGCAGTCTAGAGGATCCG 55
 RESULT 5
 AAQ11823
 ID AAQ11823 standard; DNA; 586 BP.
 AC AAQ11823;
 XX DT 24-SEP-1991 (first entry)
 XX DE Region of pUC19-SAM-K vector.
 XX XX
 XX KW S-adenosyl methionine hydrolase; ethylene biosynthesis; AdoMet;
 XX KW homoserine; MTA; ss.
 XX OS Bacteriophage T3.
 XX PH Key Location/Qualifiers
 XX FT CDS 66..524
 FT /*tag= a
 FT /label= AdoMetase
 XX PN W09109112-A.
 XX XX
 XX PD 27-JUN-1991.
 XX XX
 XX PF 12-DEC-1990; 90WO-US07175.
 XX PR 12-DEC-1989; 89US-0448095.
 XX XX (AGRI-) AGRITOPPE INC.

XX PI Ferro AJ, Bestwick RK, Brown L;
 XX XX WPI: 1991-208140/28.
 XX DR P-PSDB; AAR12520.
 XX XX
 XX PT Control of ethylene biosynthesis in plants - by introducing gene
 XX PT which encodes S-adenosyl methionine hydrolase activity
 XX XX
 XX PS Disclosure; Fig 6; 42pp; English.
 XX XX
 XX CC The sequence is a fragment of vector pUC19-SAM-K for expression of
 XX CC the AdoMetase gene. The gene is used to control ethylene biosyn-
 XX CC thesis in plants. The enzyme hydrolyses S-adenosylmethionine to
 XX CC homoserine and 5'-methylthioadenosine (MTA). The presence of the
 XX CC gene and expression of the enzyme in plants lowers levels of
 XX CC AdoMet, the sole precursor for ethylene biosynthesis, and its
 XX CC reduced availability causes a corresponding decrease in ethylene
 XX CC biosynthesis. The hydrolysis of AdoMet by the enzyme generates
 XX CC MTA which is an inhibitor of 1-aminocyclopropane-1-carboxylic
 XX CC acid (ACC)synthase, a principle enzyme in the biosynthesis of
 XX CC ethylene. This results in fruit, vegetables and flowers which
 XX CC have improved shelf life and preservation qualities.
 XX CC See also AAQ11822.
 XX XX
 XX SQ Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;
 Query Match 0.9%; Score 37; DB 12; Length 586;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AAGCTTCATGCGCTGCAGTGCAGTCTAGAGGATCCG 37
 |||||
 DB 24 AAGCTTCATGCGCTGCAGTGCAGTCTAGAGGATCCG 60
 RESULT 6
 AAQ78008
 ID AAQ78008 standard; DNA; 586 BP.
 AC AAQ78008;
 XX DT 05-JUL-1995 (first entry)
 XX DE Bacteriophage T3 SAM-K modified S-adenosylmethionine hydrolase.
 XX XX
 XX KW Tomato E8 promoter; transgenic fruit; ethylene production inhibition;
 XX KW S-adenosylmethionine hydrolase; delayed fruit ripening; ds.
 XX OS Bacteriophage T3.
 XX PH Key Location/Qualifiers
 XX FT CDS 66..524
 FT /*tag= a
 FT W09424294-A.
 XX PN
 XX PD 27-OCT-1994.
 XX XX
 XX PF 08-APR-1994; 94WO-US03886.
 XX PR 09-APR-1993; 93US-0046583.
 XX XX (EPIT-) EPITOPE INC.
 XX XX Bestwick RK, Ferro AJ;
 XX XX WPI: 1994-341873/42.
 XX DR P-PSDB; AAR74824.
 XX XX Transgenic fruit-bearing plants, esp. tomato - contg. exogenous
 XX PT gene under control of E8 promoter
 XX XX

PS Claim 18; Fig 11; 59pp; English.

XX AA078007 the tomato E8 gene promoter, and AA078008 which encodes
 CC AAR74824 the bacteriophage T3 derived SAM-K modified
 CC S-adenosylmethionine hydrolase (AdoMetase), were used in the construction
 CC of a vector. The vector was used to produce transgenic tomato plants
 CC which expressed AdoMetase under the translational control of the E8
 CC promoter. AdoMetase is an inhibitor of ethylene production, and
 CC therefore an inhibitor of fruit ripening.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 15; Length 586;

Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCG 37
 |||||
 DB 24 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCG 60

RESULT 7

AA087073

ID AA087073 standard; DNA; 586 BP.

XX AA087073;

XX 20-NOV-1995 (first entry)

XX Bacteriophage T3 S-adenosylmethionine-hydrolase gene.

XX Bacteriophage T3; phase T3; S-adenosylmethionine-hydrolase;

XX AdoMetase; SAM-hydrolase; enzyme; senescence; ethylene; hormone;

XX shelf-life; preservation; ripening; fruit; flower;

XX plasmid pUC19SAM-K; ss.

XX Bacteriophage T3.

XX Key Location/Qualifiers

XX cds 66..522

XX /*tag= a

XX US5416250-A.

XX 16-MAY-1995.

XX 12-DEC-1989; 89US-0448095.

XX 12-DEC-1989; 89US-0448095.

XX 12-DEC-1990; 90US-0613858.

XX 08-JUN-1994; 94US-0255833.

XX (AGRI-) AGRITOP INC.

XX Bestwick RK, Brown LR, Ferro AJ;

XX WPI; 1995-193491/25.

XX P-PSDB; AAR75212.

XX Reducing ethylene biosynthesis in plants - by introducing DNA
 PT encoding a S-adenosylmethionine-hydrolase enzyme into a plant host.

XX Disclosure; Fig. 6; 16pp; English.

XX This DNA encodes an S-adenosylmethionine-hydrolase enzyme which

XX hydrolyzes SAM to homoserine and 5'-methylthioadenosine. A

XX transgenic plant containing this sequence displays reduced ethylene

XX biosynthesis resulting in fruits, vegetables and flowers with

XX improved shelf life and preservation qualities.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 16; Length 586;

Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCG 37

|||||

DB 24 AAGCTTGATCCCTGCAGGTCGACTCTAGAGGATCCG 60

RESULT 8

AAT09985

ID AAT09985 standard; cDNA to mRNA; 586 BP.

XX AAT09985;

XX 11-AUG-1996 (first entry)

XX S-adenosylmethionine-hydrolase gene from pUC19-SAM-K.

XX S-adenosylmethionine-hydrolase; phase T3; chimeric gene;

XX tomato; E4 promoter; E8 promoter; fruit; transgenic plant;

XX ripening; ethylene; tissue-specific gene expression;

XX crop improvement; ds.

XX Bacteriophage T3 (ATCC 11303-B3).

XX Key Location/Qualifiers

XX misc_feature 55..83

XX /*tag= a

XX /note= "Region modified by linker to contain a

XX eukaryotic translation initiation site"

XX CDS 66..524

XX /*tag= b

XX /product= S-adenosylmethionine-hydrolase

XX WO9535387-A1.

XX 28-DEC-1995.

XX 27-OCT-1994; 94WO-US12364.

XX 17-JUN-1994; 94US-0261677.

XX (EPIT-) EPITOPE INC.

XX Bestwick RK, Ferro AJ;

XX WPI; 1996-058424/06.

XX P-PSDB; AAR88611.

XX Transgenic plants with modified ripening phenotype - contg. a
 PT chimeric gene contg. a sequence which encodes a prod. which reduces
 PT ethylene biosynthesis

XX Example 1; Page 88-89; 139pp; English.

XX The sequence represents a modified S-adenosylmethionine-hydrolase
 CC (SAM) gene isolated from phase T3 and cloned in plasmid pUC19-SAM-K
 CC (plasmid PAG-111). A synthetic double-stranded oligonucleotide
 CC linker has been used to introduce a eukaryotic translation
 CC initiation site for expression in transgenic plants. The SAM gene,
 CC which is involved in ethylene biosynthesis and fruit ripening, may
 CC be inserted in a vector under the control of a plant ripening- or
 CC ethylene biosynthesis-inducible promoter (e.g. an E4 or E8
 CC promoter, as in AAT09973) and introduced into e.g. tomato to produce
 CC a transgenic plant bearing fruit with altered fruit ripening
 CC phenotype, resulting from an initial burst of ethylene production
 CC and a subsequent reduction in ethylene biosynthesis.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 17; Length 586;

Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 37
 ID 1 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 37
 Db 24 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 60

RESULT 9

AAT09964
 ID AAT09964 standard; cDNA to mRNA; 586 BP.
 AC AAT09964;
 XX
 DT 13-MAY-1996 (first entry)
 XX
 DE S-adenosylmethionine hydrolase SAM-K cDNA.
 XX
 KW S-adenosylmethionine hydrolase; AdoMet hydrolase; AdoMetase; SAM-K;
 KW S-adenosyl cleaving enzyme; SAMase; raspberry; Rubus idaeus;
 KW strawberry; Fragaria ananassa; transgenic plant; fruit rot;
 KW disease resistance; fungus resistance; virus resistance; ethylene;
 KW ripening; crop improvement; ds.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 66..523
 FT /*tag= a
 XX
 PN W09535388-A1.
 XX
 XX 28-DEC-1995.
 PD
 XX 16-JUN-1995; 95WO-US07753.
 PF
 XX 03-FEB-1995; 95US-0384556.
 PR
 PR 17-JUN-1994; 94US-0263900.
 XX
 PA (EPIT-) EPIPOPE INC.
 XX
 PI Bestwick RK, Ferro AJ, Mathews HV;
 XX
 DR WPI: 1996-058425/06.
 DR P-PSDB: AAR88601.
 XX
 PT Transgenic red raspberry and strawberry plants - produce fruit with
 PT increased sugar content, increased fungal or viral resistance, or
 PT reduced ethylene biosynthesis
 XX
 PS Disclosure; Page 42-43; 85pp; English.

XX The coding sequence of the phage T3 S-adenosylmethionine hydrolase
 CC gene was altered (AAT09964) so that the encoded enzyme, designated
 CC SAM-K (AAR88601), had an isoleucine to valine substn. at position
 CC 2. This highly conservative change increased the translation
 CC efficiency of the gene in plants. Transgenic raspberry and
 CC strawberry plants carrying the altered gene (see also AAT09967) were
 CC obtd. showing a high-efficiency transformation and regeneration system.
 CC They showed reduced levels of ethylene biosynthesis in leaf tissue
 CC and fruit.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 17; Length 586;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 37
 ID 1 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 37
 Db 24 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 60

RESULT 10

AAT58564

ID AAT58564 standard; DNA; 586 BP.
 AC AAT58564;
 XX
 DT 17-SEP-1997 (first entry)
 XX
 DE DNA sequence of AdoMetase gene with modified 5' end.
 XX
 KW AdoMetase; bacteriophage T3; ethylene biosynthesis; transformation;
 KW S-adenosylmethionine hydrolase; hydrolysis; S-adenosylmethionine;
 KW AdoMet; homoserine; 5'-methylthioadenosine; MTA; ornamental plant;
 KW 1-aminocyclopropane-1-carboxylic acid; ACC; inhibitor; ACC synthase;
 KW transgenic plant; fruit; vegetable; flower; shelf life; preservation;
 KW food; ss.
 XX
 OS Bacteriophage T3.
 XX
 FH Key Location/Qualifiers
 FT CDS 66..524
 FT /*tag= a
 FT /product= AdoMetase
 XX
 PN US5589623-A.
 XX
 XX 31-DEC-1996.
 PD
 XX 12-DEC-1989; 89US-0448095.
 PF
 XX 12-DEC-1990; 90US-0613858.
 PR
 PR 12-DEC-1989; 89US-0448095.
 PR 08-JUN-1994; 94US-0255833.
 PR 20-DEC-1994; 94US-0360974.
 XX
 PA (AGRI-) AGRITPOPE INC.
 XX
 PI Bestwick RK, Brown LR, Ferro AJ;
 XX
 DR WPI: 1997-076910/07.
 DR P-PSDB: AAW11594.
 XX
 PT Control of ethylene biosynthesis in plants using vector encoding
 PT S-adenosyl:methionine hydrolase - used to produce fruit, flowers
 PT etc. of increased shelf life and preservation quality
 XX
 PS Example 3; Fig 6; 16pp; English.

XX This sequence represents the AdoMetase gene derived from bacteriophage
 CC T3. This sequence may be used in the method of the invention to control
 CC ethylene biosynthesis in plants. This is done by transforming plant
 CC cells with a vector which contains a DNA sequence (I) allowing genetic
 CC selection in plant cells and a DNA sequence (II) encoding S-adenosyl-
 CC methionine hydrolase (AdoMetase) which hydrolyses S-adenosylmethionine
 CC (AdoMet) to homoserine and 5'-methylthioadenosine (MTA). Both (I) and
 CC (II) are flanked by regulatory elements that allow expression in plant
 CC cells. AdoMetase reduces levels of AdoMet, the essential precursor of
 CC ethylene (via 1-aminocyclopropane-1-carboxylic acid (ACC)) and MTA is
 CC an inhibitor of ACC synthase, the major enzyme involved in ethylene
 CC synthesis. Transgenic plants containing this construct produce fruits,
 CC vegetables and flowers of improved shelf life and preservation qualities.
 CC A wide range of food and ornamental plants can be modified. The
 CC transgenic plants are able to control ethylene production under
 CC restricted conditions. This sequence has been modified around the
 CC initiation codon to produce an optimised kozak sequence.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 18; Length 586;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 37
 ID 1 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 37
 Db 24 AAGCTTCATGCCCTGCAGGTGCGACTCTAGAGGATCCG 60

Db 24 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCG 60

RESULT 11
AA000468
ID AAX00468 standard; cDNA to mRNA; 586 BP.
XX AC AAX00468;
XX DT 31-WAR-1999 (first entry)
XX DE S-adenosylmethionine hydrolase modified gene.
XX KW Ethylene; plant hormone; regulator; metabolism; senescence; E4; E8;
XX KW S-adenosylmethionine hydrolase; SAMase; promoter; fruit ripening;
XX storage; ds.
XX OS Bacteriophage t3.
XX OS Synthetic.
XX FH Key Location/Qualifiers
XX CDS 66..524
XX FT /*tag= a
XX PN US5859330-A.
XX PD 12-JAN-1999.
XX PF 27-OCT-1994; 94US-0331355.
XX PR 27-OCT-1994; 94US-0331355.
XX PR 12-DEC-1989; 89US-0448095.
XX PR 12-DEC-1990; 90US-0613858.
XX PR 09-APR-1993; 93US-0046583.
XX PR 08-JUN-1994; 94US-0255833.
XX PR 17-JUN-1994; 94US-0261677.
XX PA (EPIT-) EPITOPE INC.
XX PI Bestwick RK, Ferro AJ;
XX WPI: 1999-120030/10.
XX P-PSDB; AAW30604.
XX Transgenic plants containing S-adenosylmethionine hydrolase gene -
PT useful for delaying fruit ripening and lengthening storage periods
XX Example 1; Fig 11; 72pp; English.
XX The present invention describes a transgenic fruit-bearing plant
CC containing a DNA sequence encoding S-adenosylmethionine hydrolase
CC (SAMase) linked to a promoter selected from a group consisting of: (a)
CC a tomato E4 promoter; (b) a tomato E8 promoter; and (c) an avocado
CC cellulase promoter; (d) a tomato E4 promoter/SAMase construct.
CC Expression from the promoter is induced during fruit ripening,
CC optionally by ethylene synthesised in the fruit. The SAMase transgene
CC produces fruit with reduced ethylene synthesis, delaying fruit ripening
CC and lengthening storage periods. The present sequence represents a
CC modified SAMase gene from the present invention.
XX SQ Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;
Query Match 0.9%; Score 37; DB 20; Length 586;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCG 37
AA
DB 24 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCG 60
RESULT 12
AAA27831/c

AAA27831 standard; DNA; 3796 BP.
AAA27831;
12-SEP-2000 (first entry)
Vector plasmid pCMV-MC1.
North American PRRS virus; Nidovirales virus; pig; swine; vaccine;
pCMV-S-P129; ss.
Chimeric - Porcine reproductive and respiratory syndrome virus.
Chimeric - Human cytomegalovirus.
EP1018557-A2.
12-JUL-2000.
25-NOV-1999; 99EP-0309409.
22-DEC-1998; 98US-0113345.
(PFIZ) PFIZER PROD INC.
Calvert JG, Welch SW, Sheppard MG;
WPI: 2000-444364/39.
New polynucleotide encoding an infectious RNA molecule of a North
American porcine reproductive and respiratory syndrome virus for use as
a vaccine in protecting swine and other animals from infection by a
pathogen -
Example 4; Page 43-44; 53pp; English.
The present sequence is that of eukaryotic expression vector
plasmid pCMV-MC1. The plasmid is derived from pCMVbeta by
replacing the LacZ coding sequence with a linker containing
multiple restriction sites. Modification of the human
cytomegalovirus (CMV) immediate early promoter was accomplished by
substituting the sequence between SacI and the 2nd NciI site of
pCMV-MC1 with a synthetic linker. North American porcine
reproductive and respiratory syndrome (PRRS) virus P129A cDNA (see
AAA27809) was cloned into the modified vector to create pCMV-S-P129
(ATCC 203489), which was used to demonstrate cellular expression of
PRRS virus by direct transfection of cDNA into cells. The
invention relates to polynucleotide molecules, plasmids, viral
vectors and transfected host cells that comprise North American
PRRS DNA. It also relates to polynucleotide molecules, viral
vectors and transfected host cells encoding a genetically modified
North American PRRS virus that is disabled in its ability to cause
PRRS, or which encodes 1 or more heterologous antigenic epitopes,
for use as a vaccine.
SQ Sequence 3796 BP; 953 A; 934 C; 956 G; 953 T; 0 other;
Query Match 0.9%; Score 37; DB 21; Length 3796;
Best Local Similarity 100.0%; Pred. No. 3.5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCG 37
AA
DB 1167 AAGCTTGATCGCTGCAGGTGCTAGAGGATCCG 1131
RESULT 13
AAT15287
ID AAT15287 standard; cDNA; 4910 BP.
XX AC AAT15287;
XX DT 27-OCT-1996 (first entry)
XX

DE prEP78/52 hybrid adeno virus-adeno-associated virus cDNA.
XX
KW Adeno virus; adeno-associated virus; hybrid virus vector;
KW reporter gene; transgene; somatic gene therapy; gene transfer;
KW plasmid prEP78/52; familial hypercholesterolemia; cystic fibrosis;
KW ds.
XX
OS Adeno virus; Adeno-associated virus.
XX
PN W09613598-A2.
XX
XX 09-MAY-1996.
XX
PF 27-OCT-1995; 95WO-US14018.
XX
PR 28-OCT-1994; 94US-0331384.
XX
XX (UYPE-) UNIV PENNSYLVANIA.
PA
PI Fisher KJ, Kelley WM, Wilson JM;
XX
XX WPI; 1996-239504/24.
DR
XX Hybrid adenovirus-adeno-associated virus for gene therapy -
PT comprises adenovirus portion, 5' and 3' ITR sequences from AAV and a
PT selected transgene, e.g. CFTR or LDL
XX
PS Disclosure; Page 57-61; 91pp; English.
XX
XX This cDNA encodes recombinant plasmid prEP78/52, a trans-acting
CC plasmid containing the AAV sequences that encode rep 78 kD and
CC 52 kD proteins under the control of the AAV P5 promoter. The
CC plasmid also contains an SV40 polyadenylation signal. This cDNA
CC is used in the construction of a hybrid AAV-AV virus vector. The
CC hybrid virus may be used in the delivery and stable integration
CC of a selected gene into the chromosome of a target cell, i.e. gene
CC therapy of cystic fibrosis and familial hypercholesterolaemia.
XX
SQ Sequence 4910 BP; 1161 A; 1294 C; 1215 G; 1240 T; 0 other;
Query Match 0.9%; Score 37; DB 17; Length 4910;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 399 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 435
RESULT 14
AAV02043/c
ID AAV02043 standard; DNA; 6295 BP.
XX
XX AAV02043;
XX
DT 08-JUN-1998 (first entry)
XX
DE Plasmid pWRG3196 encoding murine interleukin-12.
XX
KW Interleukin-12; IL-12; cytokine; growth factor; mouse1
KW plasmid pWRG3196; cancer; tumour; metastasis; gene therapy; ds;
KW cyclic; circular.
XX
OS Chimeric - Mus musculus.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Bos taurus.
OS Chimeric - Rhesus macaque polyoma virus
OS Chimeric - encephalomyocarditis virus.
XX
XX Key Location/Qualifiers
FH 955..1675
FT CDS /*tag= a
FT /product= p35 subunit

FT intron /note= "contains an intron"
FT 1261..1333
FT /*tag= b
FT CDS 2377..3384
FT /*tag= c
FT /product= p40 subunit
XX
PN W09746263-A1.
XX
XX 11-DEC-1997.
PD
XX 04-JUN-1997; 97WO-US09591.
XX
XX 05-JUN-1996; 96US-0659206.
PR
XX (AURA-) AURAGEN INC.
PA
XX Rakhmillevich AL, Yang N;
PI
XX WPI; 1998-041898/04.
DR
XX P-PSDB; AAW44004-05.
DR
XX Interleukin-12 gene therapy of tumours - comprises delivering
PT construct containing promoter and sequences encoding interleukin-12
PT p35 and p40 subunit(s) to target cells in vivo
XX
PS Claim 8; Page 33-38; 50pp; English.
XX
XX Plasmid pWRG3196 is a bicistronic plasmid encoding both subunits,
CC i.e. p35 (see AAW44004) and p40 (see AAW44005), of murine interleukin-12
CC (IL-12). The p35 and p40 subunits were cloned from a mouse spleen
CC cDNA library. The vector contains a single cytomegalovirus promoter,
CC an SV40 splicing donor/splicing acceptor, and bovine growth hormone
CC polyA signal. Between the p35 and p40 genes is an internal
CC ribosome entry site element (IRES) cloned from encephalomyocarditis
CC virus. The backbone of the plasmid is from pUC19. Plasmid
CC pWRG3139 induces about half the expression of IL-12 as the tandem
CC vector pWRG3169 (see AAV02042) in vivo and in vitro. A novel method
CC of treating tumours in a mammal involves delivering copies of an
CC expressible foreign genetic construct, especially pWRG3169 or
CC pWRG3196, comprising a promoter operative in the mammalian
CC epidermal cells and DNA sequences encoding p35 and p40 subunits of
CC IL-12 to target cells in vivo. Delivery of the construct allows
CC IL-12 expression for treatment of solid, metastatic or disseminated
CC tumours, and regression of established tumours. The treatment is
CC effective even when the genetic construct is delivered to a site
CC distant from the tumour.
XX
SQ Sequence 6295 BP; 1610 A; 1610 C; 1568 G; 1507 T; 0 other;
Query Match 0.9%; Score 37; DB 19; Length 6295;
Best Local Similarity 100.0%; Pred. No. 3.4e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 3666 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 3630
RESULT 15
AAD10238/c
ID AAD10238 standard; DNA; 7164 BP.
XX
XX AAD10238;
XX
XX 24-SEP-2001 (first entry)
DT
XX Commercial plasmid vector pCMVbeta.
XX
XX Plasmid; cyclic; circular; replicon; exogenous gene; marker gene;
KW transcription termination; immunostimulatory sequence; ISS; antiviral;
KW non-essential nucleotide; molecular biology application; gene therapy;
KW DNA vaccine; cloning; gene expression; in vitro protein production;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 22:35:30 ; Search time 178 Seconds
(without alignments)
7089.746 Million cell updates/sec

Title: US-09-926-163B-1

Perfect score: 4115

Sequence: 1 aagcttgatcgctgcaggt.....tccggcgggcctgaagctt 4115

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	0.9	586	1 US-08-046-583-10	Sequence 10, Appl
2	37	0.9	586	1 US-08-384-556A-1	Sequence 1, Appl
3	37	0.9	586	2 US-08-331-355A-22	Sequence 22, Appl
4	37	0.9	586	5 PCT-US94-12364-22	Sequence 22, Appl
5	37	0.9	586	5 PCT-US95-07753-1	Sequence 1, Appl
6	37	0.9	4910	2 US-08-331-384-2	Sequence 2, Appl
7	37	0.9	4910	2 US-08-836-087-2	Sequence 2, Appl
8	37	0.9	4910	4 US-09-246-320-2	Sequence 2, Appl
9	37	0.9	4910	4 US-09-546-738-2	Sequence 2, Appl
10	37	0.9	6295	2 US-08-659-206A-4	Sequence 4, Appl
11	37	0.9	7897	3 US-08-836-022A-1	Sequence 4, Appl
12	37	0.9	7897	4 US-09-427-048A-1	Sequence 1, Appl
13	37	0.9	8540	4 US-08-487-283A-4	Sequence 4, Appl
14	37	0.9	8540	5 PCT-US96-05811A-12	Sequence 12, Appl
15	37	0.9	8932	2 US-08-252-493C-8	Sequence 8, Appl
16	37	0.9	8932	3 US-09-276-197-8	Sequence 8, Appl
17	37	0.9	9972	3 US-08-836-022A-3	Sequence 3, Appl
18	37	0.9	9972	4 US-09-427-048A-3	Sequence 3, Appl
19	37	0.9	19307	3 US-08-836-022A-10	Sequence 10, Appl
20	37	0.9	19307	4 US-09-427-048A-10	Sequence 10, Appl
21	36	0.9	40	2 US-08-660-295A-4	Sequence 4, Appl
22	36	0.9	50	2 US-08-693-302-3	Sequence 3, Appl
23	36	0.9	50	4 US-09-099-466-3	Sequence 3, Appl
24	36	0.9	57	1 US-08-089-910-19	Sequence 19, Appl
25	36	0.9	57	1 US-08-089-910-22	Sequence 22, Appl
26	36	0.9	57	4 US-08-400-864-4	Sequence 4, Appl
27	36	0.9	57	4 US-09-116-492A-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1

US-08-046-583-10
; Sequence 10, Application US/08046583
; Patent No. 5723746
; GENERAL INFORMATION:
; APPLICANT: Bestwick, Richard K
; APPLICANT: Ferro, Adolph J
; TITLE OF INVENTION: Reduced Ethylene Synthesis and Delayed
; TITLE OF INVENTION: Fruit Ripening in Transgenic Tomatoes Expressing
; TITLE OF INVENTION: S-Adenosylmethionine Hydrolase
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/046,583
; FILING DATE: 09-APR-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Fig. 11, pUC19-SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
US-08-046-583-10

Sequence 1, Appl
Patent No. 5166321
Sequence 66, Appl
Sequence 8, Appl
Sequence 5, Appl
Sequence 8, Appl
Patent No. 5304637
Patent No. 5166321
Sequence 73, Appl
Sequence 73, Appl
Sequence 73, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 1, Appl
Patent No. 5231168
Sequence 6, Appl
Sequence 7, Appl
Sequence 11, Appl

Query Match 0.9%; Score 37; DB 1; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCGCTGCAGGTGCTAGAGGATCCG 37
|||||
Db 24 AAGCTTCATGCGCTGCAGGTGCTAGAGGATCCG 60

RESULT 2

US-08-384-556A-1
; Sequence 1, Application US/08384556A
; Patent No. 5750870
; GENERAL INFORMATION:
; APPLICANT: Mathews, Helena V
; APPLICANT: Bestwick, Richard K
; TITLE OF INVENTION: Plant Genetic Transformation
; TITLE OF INVENTION: Methods and Transgenic Plants
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,556A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/263,900
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0010.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; US-08-384-556A-1

Query Match 0.9%; Score 37; DB 1; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCGCTGCAGGTGCTAGAGGATCCG 37
|||||
Db 24 AAGCTTCATGCGCTGCAGGTGCTAGAGGATCCG 60

RESULT 3

US-08-331-355A-22

; Sequence 22, Application US/08331355A
; Patent No. 5859330
; GENERAL INFORMATION:
; APPLICANT: Bestwick, Richard K
; APPLICANT: Ferro, Adolph J
; TITLE OF INVENTION: Regulated Expression of Heterologous
; TITLE OF INVENTION: Genes in Plants and Transgenic Fruit
; TITLE OF INVENTION: with a Modified Ripening Phenotype
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,355A
; FILING DATE: 27-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/261,677
; FILING DATE: 17-JUN-1994
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/046,583
; FILING DATE: 09-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/255,833
; FILING DATE: 08-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,858
; FILING DATE: 12-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 448,095
; FILING DATE: 12-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Evans, Susan T.
; REGISTRATION NUMBER: 38,443
; REFERENCE/DOCKET NUMBER: 4257-0011.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Fig. 11, pUC19-SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; US-08-331-355A-22

Query Match 0.9%; Score 37; DB 2; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCGCTGCAGGTGCTAGAGGATCCG 37
|||||
Db 24 AAGCTTCATGCGCTGCAGGTGCTAGAGGATCCG 60

RESULT 4

PCT-US94-12364-22
; Sequence 22, Application PC/TUS9412364
; GENERAL INFORMATION:
; APPLICANT: Regulated Expression of Heterologous
; TITLE OF INVENTION: Genes in Plants and Transgenic Fruit
; TITLE OF INVENTION: with a Modified Ripening Phenotype
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12364
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,677
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0011.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Fig. 11, pUC19-SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; PCT-US94-12364-22

Query Match 0.9%; Score 37; DB 5; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGATGCTGCTGAGTCTAGAGGATCCG 37
|||||
Db 24 AAGCTTGATGCTGCTGAGTCTAGAGGATCCG 60

RESULT 5
PCT-US95-07753-1
; Sequence 1, Application PC/TUS9507753
; GENERAL INFORMATION:
; APPLICANT: Plant Genetic Transformation
; TITLE OF INVENTION: Methods and Transgenic Plants
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07753
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/384,556
; FILING DATE: 03-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/263,900
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0010.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; PCT-US95-07753-1

Query Match 0.9%; Score 37; DB 5; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGATGCTGCTGAGTCTAGAGGATCCG 37
|||||
Db 24 AAGCTTGATGCTGCTGAGTCTAGAGGATCCG 60

RESULT 6
US-08-331-384-2
; Sequence 2, Application US/08331384
; Patent No. 5856152
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and Methods
; TITLE OF INVENTION: of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,384

Query Match 0.9%; Score 37; DB 2; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 399 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 435

RESULT 8
US-09-246-320-2
; Sequence 2, Application US/09246320
; Patent No. 6251677
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/246.320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836.087
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.007PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-09-246-320-2

Query Match 0.9%; Score 37; DB 2; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 399 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 435

RESULT 7
US-08-836-087-2
; Sequence 2, Application US/08836087
; Patent No. 5871982
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836.087
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/331.384
; FILING DATE: 28-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.007PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-08-836-087-2

Query Match 0.9%; Score 37; DB 2; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 399 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 435

RESULT 8
US-09-246-320-2
; Sequence 2, Application US/09246320
; Patent No. 6251677
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/246.320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836.087
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.007PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-09-246-320-2

Query Match 0.9%; Score 37; DB 4; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 399 AAGCTTCATGCTGCAGGTCGACTCTAGAGGATCCG 435

RESULT 9
US-09-546-738-2
; Sequence 2, Application US/09546738
; Patent No. 6387368
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.

;; Kelley, William M.
;; Fisher, Krishna J.
;; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
;; Methods of Use Thereof
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Howson and Howson
;; STREET: Spring House Corporate Cntr, PO Box 457
;; CITY: Spring House
;; STATE: Pennsylvania
;; COUNTRY: USA
;; ZIP: 19477
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/546,738
;; FILING DATE: 11-Apr-2000
;; CLASSIFICATION: <Unknown>
;; PRIORITY APPLICATION DATA:
;; APPLICATION NUMBER: 09/246,320
;; FILING DATE: <Unknown>
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bak, Mary E.
;; REGISTRATION NUMBER: 31,215
;; REFERENCE/DOCKET NUMBER: GNVPN.007PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-540-9200
;; TELEFAX: 215-540-5818
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4910 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;; MOLECULE TYPE: CDNA
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-546-738-2

Query Match 0.9%; Score 37; DB 4; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.le-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 399 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 435

RESULT 10
US-08-659-206A-4/c
; Sequence 4, Application US/08659206A
; Patent No. 5922685
; GENERAL INFORMATION:
; APPLICANT: Rakhmievich, Alexander
; TITLE OF INVENTION: IL-12 Gene Therapy of Tumors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,206A
; FILING DATE:

;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seay, Nicholas J.
;; REGISTRATION NUMBER: 27386
;; REFERENCE/DOCKET NUMBER: 110229.91144
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 608-251-5000
;; TELEFAX: 608-251-9166
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6295 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: circular
;; MOLECULE TYPE: other nucleic acid
;; DESCRIPTION: /desc = "plasmid pWRG3196"
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: join(955..1260, 1334..1675)
;; OTHER INFORMATION: /product= "p35 gene product"
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2377..3384
;; OTHER INFORMATION: /product= "p40 gene product"
;; US-08-659-206A-4
Query Match 0.9%; Score 37; DB 2; Length 6295;
Best Local Similarity 100.0%; Pred. No. 2.le-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 3666 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 3630
RESULT 11
US-08-836-022A-1
; Sequence 1, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Chen, Shu-Jen
; APPLICANT: Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,022A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/331,381
; FILING DATE: 28-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN.008PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7897 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-836-022A-1

Query Match 0.9%; Score 37; DB 3; Length 7897;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGTCTAGAGGATCCG 37
Db 609 AAGCTTGATCGCTGCAGTCTAGAGGATCCG 645

RESULT 12
US-09-427-048A-1
Sequence 1, Application US/09427048A
Patent No. 6203975

GENERAL INFORMATION:
APPLICANT: Trustees of the University of Pennsylvania

Wilson, James M.
Fisher, Krishna J.
Chen, Shu-Jen

Weitzman, Matthew

TITLE OF INVENTION: Improved Adenovirus Virus and
Methods of Use Thereof

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P O Box 457
CITY: Spring House

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19477

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/427,048A

FILING DATE: 21-Oct-1999

CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/836,022

FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Bak, Mary E.

REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: GNPVN.008PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-540-9200

TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7897 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: cDNA

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-427-048A-1

Query Match 0.9%; Score 37; DB 4; Length 7897;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGTCTAGAGGATCCG 37
Db 1371 AAGCTTGATCGCTGCAGTCTAGAGGATCCG 1335

Db 609 AAGCTTGATCGCTGCAGTCTAGAGGATCCG 645

RESULT 13

US-08-487-283A-4/c

Sequence 4, Application US/08487283A

Patent No. 6355245

GENERAL INFORMATION:

APPLICANT: Evans, Mark J.

APPLICANT: Mattis, Louis A.

APPLICANT: Mueller, Eileen Elliott

APPLICANT: Nye, Steven H.

APPLICANT: Rollins, Scott

APPLICANT: Rother, Russell P.

APPLICANT: Springhorn, Jeremy P.

APPLICANT: Squinto, Stephen P.

APPLICANT: Thomas, Thomas C.

APPLICANT: Wilkins, James A.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT

OF INFLAMMATORY DISEASES

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seth A. Fidel

STREET: 25 Science Park (Alexion)

CITY: New Haven

STATE: Connecticut

COUNTRY: USA

ZIP: 06511

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5 inch, 1.4Mb storage

COMPUTER: Macintosh Cetris 610

OPERATING SYSTEM: System 7

SOFTWARE: WordPerfect 3.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487,283A

FILING DATE: June 7, 1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/236,208

FILING DATE: 02-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Seth A. Fidel

REGISTRATION NUMBER: 38,449

REFERENCE/DOCKET NUMBER: ALX-152.1 CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: (203)776-1790

TELEFAX: (203)772-3655

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 8540 base pairs

TYPE: Nucleic Acid

STRANDEDNESS: Double

TOPOLOGY: Circular

MOLECULE TYPE: Other nucleic acid

DESCRIPTION: Apex-3P Eukaryotic

DESCRIPTION: Expression Vector

US-08-487-283A-4

Query Match 0.9%; Score 37; DB 4; Length 8540;

Best Local Similarity 100.0%; Pred. No. 2e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGTCTAGAGGATCCG 37
Db 1371 AAGCTTGATCGCTGCAGTCTAGAGGATCCG 1335

RESULT 14

PCT-US96-05611A-12/c

Sequence 12, Application PC/TUS9605611A

GENERAL INFORMATION:

APPLICANT: Mueller, John P.

APPLICANT: Lenardo, Michael J.

APPLICANT: McFarland, Henry F.

```

: APPLICANT: Matis, Louis A.
: APPLICANT: Mueller, Eileen Elliott
: APPLICANT: Nye, Steven H.
: APPLICANT: Pellfrey, Clara M.
: APPLICANT: Squinto, Stephen P.
: APPLICANT: Wilkins, James A.
: TITLE OF INVENTION: Modified Myelin Protein Molecules
: NUMBER OF SEQUENCES: 29
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Maurice M. Klee
: STREET: 1951 Burr Street
: CITY: Fairfield
: STATE: Connecticut
: COUNTRY: USA
: ZIP: 06430
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 0.8 Mb storage
: COMPUTER: Macintosh Centris 610
: OPERATING SYSTEM: System 7
: SOFTWARE: Microsoft Word 6.0.1
: CURRENT APPLICATION NUMBER: PCT/US96/05611A
: APPLICATION NUMBER: PCT/US96/05611A
: FILING DATE: 02-MAY-1995
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/431,644
: FILING DATE: May 2, 1995
: APPLICATION NUMBER: 08/431,648
: FILING DATE: May 2, 1995
: APPLICATION NUMBER: 08/482,114
: FILING DATE: June 7, 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Klee, Maurice M.
: REGISTRATION NUMBER: 30,399
: REFERENCE/DOCKET NUMBER: ALX-129
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (203) 255 1400
: TELEFAX: (203) 254 1101
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8540 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Double
: TOPOLOGY: Circular
: MOLECULE TYPE: Other nucleic acid
: DESCRIPTION: Apex-3P Eukaryotic
: EXPRESSION VECTOR
: PCT-US96-05611A-12

Query Match 0.9%; Score 37; DB 5; Length 8540;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGCATGCTGCAGGTCGACTCTAGAGGATCCG 37
Db 1371 AAGCTTGCATGCTGCAGGTCGACTCTAGAGGATCCG 1335

RESULT 15
US-08-252-493C-8/C
: Sequence 8, Application US/08252493C
: Patent No. 5891645
: GENERAL INFORMATION:
: APPLICANT: Rollins, Scott
: APPLICANT: Rother, Russell P.
: APPLICANT: Evans, Mark J.
: APPLICANT: Matis, Louis A.
: TITLE OF INVENTION: PORCINE E-SELECTIN
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Seth A. Fidel
: STREET: 25 Science Park, Box 15
: CITY: New Haven

```

```

: STATE: Connecticut
: COUNTRY: USA
: ZIP: 06511
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5 inch, 750 Kb storage
: COMPUTER: PC compatible
: OPERATING SYSTEM: DOS 6.2
: SOFTWARE: WordPerfect 6.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/252,493C
: FILING DATE: June 1, 1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Fidel, Seth A.
: REGISTRATION NUMBER: 38,449
: REFERENCE/DOCKET NUMBER: ALX-138
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (203) 776-1790
: TELEFAX: (203) 772-3655
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8932 base pairs
: TYPE: Nucleic Acid
: STRANDEDNESS: Double
: TOPOLOGY: Circular
: MOLECULE TYPE: Other nucleic acid
: DESCRIPTION: Apex-3 Eukaryotic
: EXPRESSION VECTOR
: US-08-252-493C-8

Query Match 0.9%; Score 37; DB 2; Length 8932;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGCATGCTGCAGGTCGACTCTAGAGGATCCG 37
Db 1371 AAGCTTGCATGCTGCAGGTCGACTCTAGAGGATCCG 1335

Search completed: May 19, 2003, 01:51:55
Job time : 235 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2003, 01:48:05 ; Search time 457 Seconds
(without alignments)
11615.089 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttgatgctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table: OIIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 810007 seqs, 644969091 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications,NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
c 1	37	0.9	3796	9	US-10-127-391-32
2	36	0.9	45	9	US-09-940-925A-161
3	36	0.9	57	10	US-09-785-269-1
c 4	36	0.9	59	9	US-09-939-275A-3
c 5	36	0.9	59	12	US-10-021-237-5
6	36	0.9	64	9	US-09-940-925A-73
7	36	0.9	66	10	US-09-785-269-8
8	36	0.9	71	10	US-09-380-932-3
9	36	0.9	75	9	US-09-972-985-6
10	36	0.9	75	10	US-09-955-649-6
11	36	0.9	75	10	US-09-973-013-6
c 12	36	0.9	89	9	US-10-135-807-11
c 13	36	0.9	102	9	US-10-015-219-106
c 14	36	0.9	102	10	US-09-777-564-106
15	36	0.9	103	10	US-09-179-536B-267
16	36	0.9	119	8	US-08-852-020-15
17	36	0.9	133	10	US-09-345-761-4
c 18	36	0.9	141	9	US-09-940-925A-162
c 19	36	0.9	149	12	US-10-012-211A-5
					Sequence 32, Appl
					Sequence 161, App
					Sequence 1, Appli
					Sequence 3, Appli
					Sequence 5, Appli
					Sequence 73, Appl
					Sequence 8, Appli
					Sequence 3, Appli
					Sequence 6, Appli
					Sequence 6, Appli
					Sequence 11, Appl
					Sequence 106, App
					Sequence 106, App
					Sequence 267, App
					Sequence 15, Appl
					Sequence 4, Appli
					Sequence 162, App
					Sequence 5, Appli

Sequence 49, Appl
Sequence 27, Appl
Sequence 32, Appl
Sequence 32, Appl
Sequence 27, Appl
Sequence 32, Appl
Sequence 50, Appl
Sequence 165, App
Sequence 11, Appl
Sequence 12, Appl
Sequence 4, Appli
Sequence 17, Appli
Sequence 6, Appli
Sequence 13, Appl
Sequence 4506, Ap
Sequence 14, Appl
Sequence 13, Appl
Sequence 53, Appl
Sequence 8, Appli
Sequence 26, Appl
Sequence 29, Appl
Sequence 1873, Ap
Sequence 1873, Ap

20 36 0.9 165 9 US-09-940-925A-49
c 21 36 0.9 206 9 US-10-033-297-27
c 22 36 0.9 206 9 US-10-081-806-32
c 23 36 0.9 206 9 US-10-074-328-32
c 24 36 0.9 206 9 US-09-940-244-27
c 25 36 0.9 206 9 US-09-940-925A-32
c 26 36 0.9 206 9 US-09-940-925A-50
c 27 36 0.9 228 9 US-09-940-925A-165
c 28 36 0.9 293 9 US-10-132-561-11
c 29 36 0.9 293 9 US-10-132-561-12
c 30 36 0.9 298 9 US-09-808-124A-4
c 31 36 0.9 298 10 US-09-845-157-6
c 32 36 0.9 355 9 US-09-910-009A-157
c 33 36 0.9 384 10 US-09-757-992-6
c 34 36 0.9 384 10 US-09-757-992-13
c 35 36 0.9 440 10 US-09-878-574-4506
c 36 36 0.9 458 9 US-10-132-561-14
c 37 36 0.9 458 9 US-10-132-561-14
c 38 36 0.9 459 9 US-10-132-561-13
c 39 36 0.9 498 10 US-09-894-882-53
c 40 36 0.9 507 10 US-09-894-882-8
c 41 36 0.9 560 10 US-09-772-134B-26
c 42 36 0.9 566 10 US-09-772-134B-29
c 43 36 0.9 586 9 US-10-052-154-1873
c 44 36 0.9 586 9 US-09-764-847-1873
c 45 36 0.9 586 10 US-09-764-847-1873

ALIGNMENTS

RESULT 1
US-10-127-391-32/c
; Sequence 32, Application US/10127391
; Patent No. US20020172690A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products Inc.
; TITLE OF INVENTION: AN INFECTIOUS CDNA CLONE OF NORTH AMERICAN PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME (PRRS) VIRUS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: PC10278A
; CURRENT APPLICATION NUMBER: US/10/127,391
; CURRENT FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US/09/470,661A
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 3796
; TYPE: DNA
; ORGANISM: Plasmid
; FEATURE:
; OTHER INFORMATION: Description of Plasmid: pCMV-MC1
US-10-127-391-32

Query Match 0.9%; Score 37; DB 9; Length 3796;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 1167 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCCG 1131

RESULT 2
US-09-940-925A-161
; Sequence 161, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

```

;
; PATHOGENS
;
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/940,925A
; FILING DATE: 10-Jun-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
;
; INFORMATION FOR SEQ ID NO: 161:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 161:
US-09-940-925A-161

Query Match 0.9%; Score 36; DB 9; Length 46;
Best Local Similarity 77.8%; Pred. No. 7.3e-09;
Matches 28; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACCTCTAGAGGATCC 36
Db 9 AAGCUUGAUGCCGACGAGGACUCUAGAGGAUCC 44

RESULT 3
US-09-785-269-1
; Sequence 1, Application US/09785269
; Patent No. US20020018736A1
; GENERAL INFORMATION:
; APPLICANT: IMAI, Kensaku
; KITAJIMA, Masato
;
; TITLE OF INVENTION: METHOD AND APPARATUS FOR AUTOMATICALLY
; REMOVING VECTOR UNIT IN DNA BASE SEQUENCE
;
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Staas & Halsey
; STREET: 700 Eleventh Street, N.W., Suite 500
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20001
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/785,269
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/684,674
; FILING DATE: 22-JUL-1996

```

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: Herbert, William F.
; REGISTRATION NUMBER: 31,024
; REFERENCE/DOCKET NUMBER: 862.1335/WEH
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 2024341500
; TELEFAX: 2024341501
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-785-269-1

Query Match 0.9%; Score 36; DB 10; Length 57;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACCTCTAGAGGATCC 36
Db 1 AAGCTTGATCGCTGCAGGTGACCTCTAGAGGATCC 36

RESULT 4
US-09-939-275A-3/C
; Sequence 3, Application US/09939275A
; Patent No. US20020172955A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Christopher
; APPLICANT: Boles, T. Christian
; APPLICANT: Weir, Lawrence
; APPLICANT: Dhanda, Rahul
; APPLICANT: Summers, Nevin
; TITLE OF INVENTION: Methods
; FILE REFERENCE: EXT-062CN
; CURRENT APPLICATION NUMBER: US/09/939,275A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/259,467
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: PCT/US98/09952
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: US 08/971,845
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/046,708
; PRIOR FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 59
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: template region
US-09-939-275A-3

Query Match 0.9%; Score 36; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACCTCTAGAGGATCC 36
Db 37 AAGCTTGATCGCTGCAGGTGACCTCTAGAGGATCC 2

RESULT 5
US-10-021-237-5/c
; Sequence 5, Application US/10021237
; Patent No. US20020119480A1
; GENERAL INFORMATION:
; APPLICANT: Weir, Lawrence

```


APPLICANT: Adams, Christopher
APPLICANT: Boles, T. Christian
APPLICANT: Dhanda, Rahul
APPLICANT: Kron, Stephen
TITLE OF INVENTION: Purification Devices Comprising Immobilized Capture Probes and Us
TITLE OF INVENTION: Therefor
FILE REFERENCE: EXT-070C1
CURRENT APPLICATION NUMBER: US/10/021,237
PRIORITY FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: US 09/513,381
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/121,836
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 59
TYPE: DNA
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: polylinker.
US-10-021-237-5

Query Match 0.9% Score 36; DB 12; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 36
|||||
Db 37 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 2
|||||

RESULT 6
US-09-940-925A-73
Sequence 73, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
LYAMICHEV, VICTOR I.
OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-09-940-925A-73

Query Match 0.9% Score 36; DB 9; Length 64;
Best Local Similarity 77.8%; Pred. No. 7.2e-09;
Matches 28; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 36
|||||
Db 9 AAGCUUGCAUGCCUGCAGGUGGACUCUAGAGGAUCC 44
|||||

RESULT 7
US-09-785-269-8
Sequence 8, Application US/09785269
Patent No. US20020018736A1
GENERAL INFORMATION:
APPLICANT: IMAI, Kensaku
KITAJIMA, Masato
TITLE OF INVENTION: METHOD AND APPARATUS FOR AUTOMATICALLY
REMOVING VECTOR UNIT IN DNA BASE SEQUENCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Staas & Halsey
STREET: 700 Eleventh Street, N.W., Suite 500
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/785,269
FILING DATE: 20-Feb-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/684,674
FILING DATE: 22-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Herbert, William F.
REGISTRATION NUMBER: 31,024
REFERENCE/DOCKET NUMBER: 862.1335/WFH
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2024341500
TELEFAX: 2024341501
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 66
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-785-269-8

Query Match 0.9% Score 36; DB 10; Length 66;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 36
|||||
Db 6 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 41
|||||

RESULT 8
US-09-380-932-3
Sequence 3, Application US/09380932
Patent No. US20020058250A1
GENERAL INFORMATION:
APPLICANT: FIRTH, Greg

; TITLE OF INVENTION: EXTRACTION AND UTILISATION OF VNTR ALLELES
; FILE REFERENCE: 28911/35930
; CURRENT APPLICATION NUMBER: US/09/380,932
; CURRENT FILING DATE: 1999-03-21
; PRIOR APPLICATION NUMBER: PCT/GB98/00840
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: EP 97301917.7
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 71
; TYPE: DNA
; ORGANISM: genomic DNA
US-09-380-932-3

Query Match 0.9%; Score 36; DB 10; Length 71;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 4 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 39

RESULT 9

US-09-972-985-6
; Sequence 6, Application US/09972985
; Publication No. US20030066395A1
; GENERAL INFORMATION:
; APPLICANT: Ladtant, Daniel
; APPLICANT: Karimova, Agnes
; APPLICANT: Ullmann, Agnes
; TITLE OF INVENTION: BACTERIAL MULTI-HYBRID SYSTEM AND APPLICATIONS
; FILE REFERENCE: 03495.0178
; CURRENT APPLICATION NUMBER: US/09/972,985
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: 09/203,681
; PRIOR FILING DATE: 1998-12-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; NAME/KEY: CDS
; LOCATION: (1)..(75)
US-09-972-985-6

Query Match 0.9%; Score 36; DB 9; Length 75;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 18 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 53

RESULT 10

US-09-955-649-6
; Sequence 6, Application US/09955649
; Patent No. US20020033812A1
; GENERAL INFORMATION:
; APPLICANT: Ladtant, Daniel
; APPLICANT: Karimova, Gouzel
; APPLICANT: Ullmann, Agnes
; TITLE OF INVENTION: BACTERIAL MULTI-HYBRID SYSTEM AND APPLICATIONS THEREOF
; FILE REFERENCE: 03495.0190
; CURRENT APPLICATION NUMBER: US/09/955,649
; CURRENT FILING DATE: 2000-08-01

; PRIOR APPLICATION NUMBER: PCT/IB98/02085
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; NAME/KEY: CDS
; LOCATION: (1)..(75)
US-09-955-649-6

Query Match 0.9%; Score 36; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 18 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 53

RESULT 11

US-09-973-013-6
; Sequence 6, Application US/09973013
; Patent No. US20020106783A1
; GENERAL INFORMATION:
; APPLICANT: Ladtant, Daniel
; APPLICANT: Karimova, Gouzel
; APPLICANT: Ullmann, Agnes
; TITLE OF INVENTION: BACTERIAL MULTI-HYBRID SYSTEM AND APPLICATIONS THEREOF
; FILE REFERENCE: 03495.0178-01
; CURRENT APPLICATION NUMBER: US/09/973,013
; CURRENT FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 09/203,681
; PRIOR FILING DATE: 1998-12-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; NAME/KEY: CDS
; LOCATION: (1)..(75)
US-09-973-013-6

Query Match 0.9%; Score 36; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 18 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 53

RESULT 12

US-10-135-807-11/C
; Sequence 11, Application US/10135807
; Publication No. US20030049655A1
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. US20030049655A1
; APPLICANT: Vincent, Suzanne P
; APPLICANT: Clark, Duncan Roy
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: GB 0110501.4

```

; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 89
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Modified
; OTHER INFORMATION: polylinker sequence of pTQ18NHK
US-10-135-807-11

Query Match          0.9%; Score 36; DB 9; Length 89;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 36
    |||||||
Db 89 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 54
    |||||||

RESULT 13
US-10-015-219-106/c
; Sequence 106, Application US/10015219
; Publication No. US2003008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 102
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2
; OTHER INFORMATION: n = A,T,C or G
US-10-015-219-106

Query Match          0.9%; Score 36; DB 9; Length 102;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 36
    |||||||
Db 77 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 42

RESULT 14
US-09-777-564-106/c
; Sequence 106, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 106
; LENGTH: 102
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (1)...(102)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-106

Query Match          0.9%; Score 36; DB 10; Length 102;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 36
    |||||||
Db 77 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 42

RESULT 15
US-09-179-536B-267
; Sequence 267, Application US/09179536B
; Patent No. US20020042112A1
; GENERAL INFORMATION:
; APPLICANT: Hubert K ster
; David M. Lough
; Guobing Xiang
; TITLE OF INVENTION: DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY
; NUMBER OF SEQUENCES: 320
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/179,536B
; FILING DATE: 26-Oct-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/20444
; FILING DATE: 06-NOV-1997
; APPLICATION NUMBER: 08/947,801
; FILING DATE: 08-Oct-97
; APPLICATION NUMBER: 08/933,792
; FILING DATE: 19-Sep-97
; APPLICATION NUMBER: 08/787,639
; FILING DATE: 23-Jan-97
; APPLICATION NUMBER: 08/786,988
; FILING DATE: 23-Jan-97
; APPLICATION NUMBER: 08/746,055
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/746,036
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/744,590
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/744,481
; FILING DATE: 06-No. US20020042112A1-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24736-2004B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8400
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 267:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA

```

```

;
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 267:
US-09-179-536B-267

```

```

Query Match      0.9% Score 36; DB 10; Length 103;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 AAGCTTGATCGCTGCAGGTCGACTCTAGAGGATCC 36
         |||||||
Db      21 AAGCTTGATCGCTGCAGGTCGACTCTAGAGGATCC 56

```

Search completed: May 19, 2003, 07:50:56
Job time : 461 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2003, 02:27:41 ; Search time 5382 Seconds
(without alignments)
12382.837 Million cell updates/sec

Title: US-09-926-163B-1

Perfect score: 4115

Sequence: 1 aagctgcacgcctgcaggt.....tccggcgccctgaagctt 4115

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST.*

1: em_estba.*

2: em_esthum.*

3: em_estin.*

4: em_estmu.*

5: em_estov.*

6: em_estpl.*

7: em_estro.*

8: em_htc.*

9: gb_est1.*

10: gb_est2.*

11: gb_hic.*

12: gb_est3.*

13: gb_est4.*

14: gb_est5.*

15: em_estfun.*

16: em_estom.*

17: gb_gss.*

18: em_gss_hum.*

19: em_gss_inv.*

20: em_gss_pin.*

21: em_gss_vrt.*

22: em_gss_fun.*

23: em_gss_mam.*

24: em_gss_mus.*

25: em_gss_other.*

26: em_gss_pro.*

27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	0.9	241	14	Z25777 pha-I-2R/1L
C 2	38	0.9	534	17	B60048 CIT-HSP-385
C 3	38	0.9	617	17	B70879 CIT-HSP-206
C 4	38	0.9	672	17	B70567 CIT-HSP-205
C 5	38	0.9	739	17	B64148 CIT-HSP-202
C 6	38	0.9	763	17	AQ916288 nbeb00630

C 7	37	0.9	63	17	B32178
C 8	37	0.9	123	17	B40882
C 9	37	0.9	161	17	B79682
C 10	37	0.9	205	14	BM851683
C 11	37	0.9	206	14	BM850221
C 12	37	0.9	207	17	CNS00271
C 13	37	0.9	210	17	B33347
C 14	37	0.9	218	9	AT000691
C 15	37	0.9	221	17	AQ006733
C 16	37	0.9	246	17	B31188
C 17	37	0.9	253	2	BM821080
C 18	37	0.9	255	2	BM740520
C 19	37	0.9	285	17	B31114
C 20	37	0.9	285	17	B99340
C 21	37	0.9	290	14	BM850232
C 22	37	0.9	294	17	B38698
C 23	37	0.9	296	17	AQ080154
C 24	37	0.9	301	17	AQ080152
C 25	37	0.9	304	17	B40812
C 26	37	0.9	306	17	AQ002527
C 27	37	0.9	308	17	B34609
C 28	37	0.9	309	17	B32855
C 29	37	0.9	309	17	B41250
C 30	37	0.9	313	14	BM851117
C 31	37	0.9	315	17	B32887
C 32	37	0.9	319	2	BM852732
C 33	37	0.9	321	2	BM850274
C 34	37	0.9	322	17	B31778
C 35	37	0.9	323	2	BM821456
C 36	37	0.9	324	14	BM821148
C 37	37	0.9	330	14	BM849662
C 38	37	0.9	331	2	BM852441
C 39	37	0.9	333	2	BM747927
C 40	37	0.9	334	2	BM829792
C 41	37	0.9	341	17	B30702
C 42	37	0.9	342	17	B32593
C 43	37	0.9	344	17	AQ059544
C 44	37	0.9	348	2	BM852504
C 45	37	0.9	348	17	B45032

ALIGNMENTS

RESULT 1
Z25777
LOCUS 241 bp mRNA linear EST 08-OCT-1998
DEFINITION pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

ACCESSION Z25777
VERSION Z25777.1 GI:3702537

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS 1 (bases 1 to 241)

Vinogradova, T.V., Lebedev, Y.B., Kopantzev, E.P., Wagner, L.L., Volik

S.V. Ermolaeva, O.D., Lavrentyeva, I., Monastyrskaya, G.S. and

Sverdlov, E.D.

Outward Alu-primed hncDNA library

Unpublished (1995)

COMMENT Contact: Sverdlov ED

Structure and Function of Human Genes

Shemyakin Institute of Bioorganic Chemistry

16/10 Miklukho-Maklaya, Moscow, 117871, Russia

Tel: 70953306529

Fax: 70953306538

Email: sverd@humgen.siobc.msk.su.

Location/Qualifiers

1..241

FEATURES

source

/db_xref="Homo sapiens"

/taxon:9606"

```

/clone="pha-i-2R/1LW"
/clone_lib="Outward Alu-primed hmcDNA library"
/notes="Vector: pGEM-32; Site_1: EcoRI; Site_2: BamHI; The
library was constructed as described in [Obradovic, D.,
Borodin, A.M., Kopantsev, E.P., Wagner, L.L., Volik, S.V.,
Ermolaeva, O.D., Lebedev, Y.B., Monastyrskaya, G.S.,
Sverdlov, E.D. (1993) Bioorganicheskaya khimiya, 20,
919-930]. This protocol is based on nested primer strategy
using Alu- specific primers (ALN3 and TC-65) that direct
the hmcDNA synthesis outward of Alu repeats."
BASE COUNT      72.a  64 c  44 g  55 t
ORIGIN

```

```

Query Match      0.9%; Score 39; DB 14; Length 241;
Best Local Similarity 100.0%; Pred. No. 4.1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  2 AGCTTGCATGCTCGAGTCGACTCTAGAGGATCCGGTT 40
      |||||||
Db   7 AGCTTGCATGCTCGAGTCGACTCTAGAGGATCCGGTT 45

```

```

RESULT 2
B60048/c
LOCUS      B60048      534 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-385K18.TFB CIT-HSP Homo sapiens genomic clone 385K18, DNA
sequence.
ACCESSION  B60048
VERSION    B60048.1 GI:2614766
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE 1 (bases 1 to 534)
AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: CIT-HSP-385K18.TRB
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

```

```

FEATURES             Location/Qualifiers
     source           1..534
                     /organism="Homo sapiens"
                     /db_xref="GDB:5378802"
                     /db_xref="taxon:9606"
                     /clone="385K18"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"

```

```

BASE COUNT      175 a  145 c  102 g  112 t
ORIGIN

Query Match      0.9%; Score 38; DB 17; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  1 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 38
      |||||||
Db   47 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 10

```

```

RESULT 4
B70567/c
LOCUS      B70567      672 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-2059L21.TF CIT-HSP Homo sapiens genomic clone 2059L21, DNA
sequence.
ACCESSION  B70567
VERSION    B70567.1 GI:2709791
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE 1 (bases 1 to 672)
AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden

```

```

Db   47 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 10
      |||||||

```

```

RESULT 3
B70879/c
LOCUS      B70879      617 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-2063C19.TF CIT-HSP Homo sapiens genomic clone 2063C19, DNA
sequence.
ACCESSION  B70879
VERSION    B70879.1 GI:2710103
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE 1 (bases 1 to 617)
AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: CIT-HSP-2063C19.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

```

```

FEATURES             Location/Qualifiers
     source           1..617
                     /organism="Homo sapiens"
                     /db_xref="GDB:7061446"
                     /db_xref="taxon:9606"
                     /clone="2063C19"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"

```

```

BASE COUNT      176 a  147 c  139 g  155 t
ORIGIN

```

```

Query Match      0.9%; Score 38; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  1 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 38
      |||||||
Db   47 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 10

```

```

RESULT 4
B70567/c
LOCUS      B70567      672 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-2059L21.TF CIT-HSP Homo sapiens genomic clone 2059L21, DNA
sequence.
ACCESSION  B70567
VERSION    B70567.1 GI:2709791
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE 1 (bases 1 to 672)
AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
and Venter,J.C.
TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
Building
JOURNAL   Unpublished (1997)
COMMENT   Other_GSSs: CIT-HSP-2063C19.TR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: M13-21
Class: BAC ends.

```

```

FEATURES             Location/Qualifiers
     source           1..617
                     /organism="Homo sapiens"
                     /db_xref="GDB:7061446"
                     /db_xref="taxon:9606"
                     /clone="2063C19"
                     /clone_lib="CIT-HSP"
                     /sex="Male"
                     /cell_type="Sperm"
                     /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
HindIII"

```

```

BASE COUNT      176 a  147 c  139 g  155 t
ORIGIN

```

```

Query Match      0.9%; Score 38; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  1 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 38
      |||||||
Db   47 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 10

```

```

RESULT 4
B70567/c
LOCUS      B70567      672 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-2059L21.TF CIT-HSP Homo sapiens genomic clone 2059L21, DNA
sequence.
ACCESSION  B70567
VERSION    B70567.1 GI:2709791
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE 1 (bases 1 to 672)
AUTHORS   Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden

```

K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map Building
 Unpublished (1997)
 Other_GSSs: CIT-HSP-2059L21.TR
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..672
 /organism="Homo sapiens"
 /db_xref="GDB:7060128"
 /db_xref="taxon:9606"
 /clone="2059L21"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 195 a 118 c 118 g 241 t

ORIGIN

Query Match 0.9%; Score 38; DB 17; Length 672;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCGG 38
 |||||||

Db 38 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCGG 1

RESULT 5

B64148/c

LOCUS B64148 739 bp DNA linear GSS 21-JUN-1998
 DEFINITION CIT-HSP-2021E8.TF CIT-HSP Homo sapiens genomic clone 2021E8, DNA sequence.

ACCESSION B64148

VERSION B64148.1 GI:2638138

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 739)

ADAMS, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map Building

Unpublished (1997)

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdadams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC end search page:

http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html

Seq primer: M13-21

Class: BAC ends.

FEATURES

Location/Qualifiers

source

1..739
 /organism="Homo sapiens"
 /db_xref="GDB:7045354"
 /db_xref="taxon:9606"
 /clone="2021E8"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 246 a 131 c 176 g 186 t

ORIGIN

Query Match 0.9%; Score 38; DB 17; Length 739;

Best Local Similarity 100.0%; Pred. No. 1.2e-08;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCGG 38
 |||||||

Db 61 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCGG 24

RESULT 6

AQ916288/c

LOCUS

DEFINITION AQ916288 763 bp DNA linear GSS 02-DEC-1999
 clone nb0063017f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic

ACCESSION AQ916288

VERSION AQ916288.1 GI:6512804

KEYWORDS GSS.

SOURCE Oryza sativa.

ORGANISM

Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.

1 (bases 1 to 763)

Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome

Unpublished (1998)

Contact: Wing RA

Clemson University Genomics Institute

Clemson University

100 Jordan Hall, Clemson, SC 29634, USA

Tel: 864 656 7288

Fax: 864 656 4293

Email: rwing@clemson.edu

Seq primer: TAATAGGACTCACTATAGGG

Class: BAC ends

High quality sequence start: 43

High quality sequence stop: 412.

Location/Qualifiers

FEATURES

source

1..763
 /organism="Oryza sativa"
 /strain="Japonica"
 /cultivar="Nipponbare"
 /db_xref="taxon:4530"
 /clone="nb0063017f"
 /clone_lib="CUGI Rice BAC Library (EcoRI)"
 /tissue_type="Leaf"
 /lab_host="E. coli DH10B"

/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;
 Rice is the most important food crop in the world. Half of
 the world population, especially those inhabiting highly
 populated areas of the humid tropics and subtropics, rely
 on rice as their primary source of carbohydrate.

Monocotyledonous rice is a diploid plant (2n=24) with a
 haploid genome equivalent of 431 Mbp (Arumuganathan and
 Earle, 1991). The relatively small genome of rice, three
 times larger than that of Arabidopsis, makes it suitable
 for genomic studies. In order to facilitate positional
 cloning, physical mapping and genome sequencing of rice,
 we have constructed a BAC library from Oryza sativa,
 Nipponbare variety using EcoRI as the cloning enzyme. The

library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 204 a 195 c 163 g 197 t 4 others
 ORIGIN

Query Match 0.9%; Score 38; DB 17; Length 763;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 38
 |||||||
 Db 68 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 31
 |||||||

RESULT 7
 B32178/c
 LOCUS B32178 63 bp DNA linear GSS 17-OCT-1997
 DEFINITION HS-1014-B2-E05-MF.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 789 Col=10 Row=J, DNA sequence.
 ACCESSION B32178
 VERSION B32178.1 GI:2531547
 KEYWORDS GSS
 SOURCE human.

REFERENCE 1 (bases 1 to 63)
 AUTHORS Mahairas G.G., Zackrone, K.D., Smith, T., Tipton, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.
 TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
 JOURNAL Unpublished (1997)
 COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
 University of Washington
 Seattle, WA 98195, USA
 Tel: (206) 616-8744
 Fax: (206) 685-7301
 Email: kzackrone@u.washington.edu
 Sequence Tagged Connector
 Plate: CT 789 row: J column: 10
 Class: BAC ends
 High quality sequence stop: 63.
 Location/Qualifiers
 1. .63
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Human Genomic Sperm Library C"
 /sex="M"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 14 a 18 c 18 g 13 t
 ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 63;
 Best Local Similarity 100.0%; Pred. No. 6.1e-08;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 37
 |||||||
 Db 51 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 15
 |||||||

RESULT 8
 B40882/c
 LOCUS B40882 123 bp DNA linear GSS 18-OCT-1997

DEFINITION HS-1052-B1-F02-MF.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 774 Col=3 Row=L, DNA sequence.
 ACCESSION B40882
 VERSION B40882.1 GI:2545134
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 123)
 Mahairas, G.G., Zackrone, K.D., Smith, T., Tipton, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.
 TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
 JOURNAL Unpublished (1997)
 COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
 University of Washington
 Seattle, WA 98195, USA
 Tel: (206) 616-8744
 Fax: (206) 685-7301
 Email: kzackrone@u.washington.edu
 Sequence Tagged Connector
 Plate: CT 774 row: L column: 3
 Class: BAC ends
 High quality sequence stop: 123.
 Location/Qualifiers
 1. 123
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Human Genomic Sperm Library C"
 /sex="M"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 39 a 20 c 24 g 40 t
 ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 123;
 Best Local Similarity 100.0%; Pred. No. 5.5e-08;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 37
 |||||||
 Db 49 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 13
 |||||||

RESULT 9
 B79682/c
 LOCUS B79682 161 bp DNA linear GSS 24-OCT-1998
 DEFINITION CIT-HSP-2045K8.TF CIT-HSP Homo sapiens genomic clone 2045K8, DNA sequence.
 ACCESSION B79682
 VERSION B79682.1 GI:2866705
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 161)
 Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.
 TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
 JOURNAL Unpublished (1998)
 COMMENT Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org

BASE COUNT 14 a 18 c 18 g 13 t
 ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 63;
 Best Local Similarity 100.0%; Pred. No. 6.1e-08;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 37
 |||||||
 Db 51 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCCG 15
 |||||||

RESULT 8
 B40882/c
 LOCUS B40882 123 bp DNA linear GSS 18-OCT-1997

Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

Location/Qualifiers

```
1..161
/organism="Homo sapiens"
/db_xref="GB:7054715"
/db_xref="taxon:9606"
/clone="2045K8"
/clone_lib="CIR-HSP"
/sex="Male"
/cell_type="Sperm"
/Note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"
```

BASE COUNT 47 a 42 c 37 g 35 t
 ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 161;
 Best Local Similarity 100.0%; Pred. No. 5.2e-08;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCG 37

DB 59 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCG 23

RESULT 10

BM851683/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 42 row: B column: 01

High quality sequence stop: 205.

Location/Qualifiers

1..205

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S21SNU520-42-B01"

/clone_lib="S21SNU520"

/sex="F"

/tissue_type="Stomach"

/cell_type="Floating aggregates"

/lab_host="Top10F"

/Note="Organ: Stomach; Vector: pTZ18Rpl; Site_1: EcoRI;

Site_2: NotI; The poly (A)+ RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of 14 RNA ligase and the first strand

cDNA was synthesized from oligo dt-selected mRNA by

priming with dt-tailed vector. The dt-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

BASE COUNT 39 a 59 c 63 g 44 t
 ORIGIN

Query Match

Best Local Similarity

Matches

QY

DB

RESULT 11

BM850221/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 41 row: A column: 10

High quality sequence stop: 206.

Location/Qualifiers

1..206

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S21SNU520-41-A10"

/clone_lib="S21SNU520"

/sex="F"

/tissue_type="Stomach"

/cell_type="Floating aggregates"

/lab_host="Top10F"

/Note="Organ: Stomach; Vector: pTZ18Rpl; Site_1: EcoRI;

Site_2: NotI; The poly (A)+ RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tobacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of 14 RNA ligase and the first strand

cDNA was synthesized from oligo dt-selected mRNA by

priming with dt-tailed vector. The dt-tailed vector was

adjusted to have about 60nt. The cDNA vector was

circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand

converted to a DNA strand by Okayama-Berg method. The

obtained cDNA vectors were used for transformation of

competent cells E. coli Top10F by electroporation method.

The cDNA libraries constructed by this method are full-length enriched cDNA library."

39 a 56 c 65 g 46 t

Query Match 0.9%; Score 37; DB 14; Length 206;
Best Local Similarity 100.0%; Pred. No. 5.1e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 170 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 134
|||||

RESULT 12
CNS00271/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN01A13 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL097303
VERSION AL097303.1 GI:5608914
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 207)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
source
1..207
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN01A13"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : T7"

BASE COUNT 40 a 62 c 40 g 62 t 3 others

Query Match 0.9%; Score 37; DB 17; Length 207;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 187 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 151
|||||

RESULT 13
B33347/c
LOCUS
DEFINITION HS-1017-A2-B08-MF.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 792 Col=16 Row=C, DNA sequence.
ACCESSION B33347
VERSION B33347.1 GI:2532716
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 210)
Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
Unpublished (1997)
Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu
Sequence Tagged Connector
Plate: CT 792 row: C column: 16
Class: BAC ends
High quality sequence stop: 210.
Location/Qualifiers
1..210
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 792 Col=16 Row=C"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 53 a 48 c 44 g 65 t

ORIGIN
Query Match 0.9%; Score 37; DB 17; Length 210;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 51 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 15
|||||

RESULT 14
AT000691/c
LOCUS
DEFINITION AT000691 Brassica rapa guard cell Brassica rapa subsp. pekinensis
cDNA clone DGT694, mRNA sequence.
ACCESSION AT000691
VERSION AT000691.1 GI:3414225
KEYWORDS EST.
SOURCE Brassica rapa subsp. pekinensis.
ORGANISM Brassica rapa subsp. pekinensis.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 218)
AUTHORS Kwak,J.M., Kim,S.A., Hong,S.W. and Nam,H.G.
TITLE Evaluation of 515 expressed sequence tags obtained from guard cells of Brassica campestris
JOURNAL Planta 202 (1), 9-17 (1997)
MEDLINE 97320163
COMMENT Contact: Hong-Gil Nam
Department of Life Science, Plant Molecular Genetics Laboratory
Pohang University of Science and Technology
San 31 Hyodong, Pohang Kyungbuk 790-784, Korea
Email: hgn@bric.postech.ac.kr
Submitted through BRIC(Biological Research Information Center) of Korea URL: http://bric.postech.ac.kr/.

FEATURES
source
1..218
/organism="Brassica rapa subsp. pekinensis"
/db_xref="taxon:51351"
/clone="DGT694"
/clone_lib="Brassica rapa guard cell"
/cell_type="guard cell protoplast"

BASE COUNT 60 a 42 c 29 g 87 t

ORIGIN

Query Match 0.9%; Score 37; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 217 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 181
|||||

RESULT 15

AQ006733/c
LOCUS CIT-HSP-2291D16.TF CIT-HSP Homo sapiens genomic clone 2291D16, DNA
DEFINITION 221 bp DNA linear GSS 27-JUN-1998
sequence.
ACCESSION AQ006733
VERSION AQ006733.1 GI:3083959
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 221)
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2291D16.FR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES

source Location/Qualifiers
1..221
/organism="Homo sapiens"
/db_xref="GDB:7150963"
/db_xref="taxon:9606"
/clone="2291D16"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT

48 a 63 c 48 g 62 t

ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 221;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 77 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 41
|||||

Search completed: May 19, 2003, 09:35:37
Job time : 5405 secs

THIS PAGE BLANK (U&PTO)

THIS PAGE BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 18:00:11 ; Search time 10356 Seconds
(without alignments)
11564.122 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttgatgcctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_btg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pet.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4073	99.0	4073	1	AB028937	AB028937 Gluconoba
2	955.6	23.2	1458	1	AB063188	AB063188 Gluconoba
3	955.6	23.2	1458	6	AX012443	AX012443 Sequence
4	955.6	23.2	1458	6	E50685	E50685 Gene manipu
5	351.4	8.5	10977	1	AE004660	AE004660 Pseudomon
6	327.4	8.0	10031	1	AF007800	AF007800 Pseudomon
7	292.4	7.1	332635	1	AP003005	AP003005 Mesorhizo
8	268.4	6.5	323450	1	SME591790	AL591790 Sinorhizo
9	262.2	6.4	189370	1	AF010496	AF010496 Rhodobact
10	259.2	6.3	11164	1	AE009374	AE009374 Agrobacte
11	259.2	6.3	14945	1	AE008240	AE008240 Agrobacte
12	247.4	6.0	9810	1	AF018073	AF018073 Rhodobact
13	236.8	5.8	10955	1	AE005663	AE005663 Escherich
14	236.8	5.8	168764	1	AP002569	AP002569 Escherich
15	236.2	5.7	189333	1	RME603647	AL603647 Rhizobium
16	224	5.4	4539	1	D13329	D13329 Escherichia
17	224	5.4	11277	1	AE000503	AE000503 Escherich
18	224	5.4	338534	1	ECOUW93	U14003 Escherichia
19	219.4	5.3	11047	1	AE012535	AE012535 Xanthomon
20	213.6	5.2	1509	6	AX120210	AX120210 Sequence
21	213.6	5.2	1632	6	AX065209	AX065209 Sequence
22	213.6	5.2	340000	1	AP005274	AP005274 Corynebac
23	213.6	5.2	349980	6	AX120085	AX120085 Sequence
24	210.2	5.1	10788	1	AE012076	AE012076 Xanthomon
25	210	5.1	266050	1	AL627270	AL627270 Salmonell
26	206.8	5.0	21585	1	AE008765	AE008765 Salmonell
27	203.2	4.9	25034	1	AE008844	AE008844 Salmonell
28	202.8	4.9	10594	1	AE005823	AE005823 Caulobact
29	198.4	4.8	230050	1	AL627277	AL627277 Salmonell
30	186.8	4.5	14681	1	AE013893	AE013893 Yersinia
31	186.8	4.5	199050	1	AJ414147	AJ414147 Yersinia
32	184	4.5	24336	1	AF008841	AF008841 Salmonell
33	181.8	4.4	347750	1	AP002998	AP002998 Mesorhizo
34	179	4.3	286485	1	AP002557	AP002557 Escherich
35	178	4.3	11783	1	AE005350	AE005350 Escherich
36	177.4	4.3	11258	1	AE000251	AE000251 Escherich
37	177.4	4.3	14045	1	D90797	D90797 E.coli geno
38	177.4	4.3	19465	1	D90798	D90798 E.coli geno
39	173.4	4.2	10446	1	AE000306	AE000306 Escherich
40	173.4	4.2	10449	1	AE005449	AE005449 Escherich
41	173.4	4.2	75888	1	ECOHU47	U00007 47 to 48 ce
42	173.4	4.2	268857	1	AP002560	AP002560 Escherich
43	171.8	4.2	17093	1	D90849	D90849 E.coli geno
44	170.6	4.1	244894	2	AC084324	AC084324 Mus muscu
45	167	4.1	3250	8	SCYNR073C	Z71688 S.cerevisia

ALIGNMENTS

RESULT 1	AB028937	4073 bp	DNA	linear	BCT 11-JUL-2000
LOCUS					
DEFINITION	AB028937	Gluconobacter oxydans gene for sorbitol dehydrogenase, complete cds.			
ACCESSION	AB028937				
VERSION	AB028937.1	GI:9049448			
KEYWORDS	sorbitol dehydrogenase.				
SOURCE	Gluconobacter oxydans (strain:G624)	DNA.			
ORGANISM	Gluconobacter oxydans				
REFERENCE	Bacteria: Proteobacteria; alpha subdivision; Acetobacteraceae;				
AUTHORS	1 (sites)				
	Shibata,T., Ichikawa,C., Matsura,M., Takata,Y., Noguchi,Y.,				
	Saito,Y. and Yamashita,M.				

TITLE Cloning of a gene for D-Sorbitol dehydrogenase from *Gluconobacter oxydans* G624 and expression of the gene in *Pseudomonas putida* IF03738

JOURNAL J. Biosci. Bioeng. 89, 463-468 (2000)

REFERENCE 2. (Bases 1 to 4073)

AUTHORS Shibata, T. and Saito, Y.

TITLE Direct Submission

JOURNAL Submitted (15-JUN-1999) Takashi Shibata, Fujisawa Pharmaceutical Co., Ltd., Fermentation Development Laboratories; Shinkawa-cho, Nakagawa 156, Nishikasuga-gun, Aichi 452-0915, Japan (E-mail: takashi.shibata@fujisawa.co.jp, Tel: 81-52-401-0275, Fax: 81-52-400-1380)

FEATURES

Location/Qualifiers

1. 4073

organism="Gluconobacter oxydans"

/strain="G624"

/db_xref="taxon:442"

501..1958

CDS

/codon_start=1

/transl_table=11

/product="sorbitol dehydrogenase"

/protein_id="BA099414.1"

/db_xref="gi:9049449"

/translation="MITRETLKSLPANVQAPPYDIDGKIVHFGVGNFFRAHEAFYVEQLEHAPDWAIVGVGLTSDRSKKRAEEFKADQCLYSLETAPSGKSTVRVHGALRDYLLAPADPEAVLKHVPATRIYVSMITTEGGYNINETGAFDLENAARVADLKNPKPSTVFGVYVEALRRRDAGKAFVMSCDNLRHNGNVARKAFGLYAKARDELAKWIEENATFPNGWDRITPTVSAEIAKLNAAAGLDDPLVAEDFHOWVLEDFADGRPELENKQVQWQVDWEYKIRMLNAGHVMCLFPGLVGYENVDDAIEDSELGNLKNYLKNDVPTLKAPSGMTLEGYRDSVSRFSNKMADOTLRIASDSCSKVQVFWETVRRLEFDRDLRSIAFGIATSIEMLRDERGKGYESESSEPTYGDAEMKLAKADDFESSLKLPAFDGWRDLDTSELQKIVLRLKIREKGVKAIPA"

BASE COUNT 823 a 1122 c 1259 g 869 t

ORIGIN

Query Match 99.0%; Score 4073; DB 1; Length 4073;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 4073; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 37 GGTTTTGGCAGCGTCCCTAGATTGATCGCGCTGCTGTGACCGACATGATGCTGGTGGC 96

DB 1 GGTTTTGGCAGCGTCCCTAGATTGATGCGCGCTGCTGTGACCGACATGATGCTGGTGGC 60

QY 97 ACGTGCCATTGCGACGGGGCGTGGACCGGGAACACAGCGCTGCTGCTTTGTACAAGG 156

DB 61 ACGTGCCATTGCGACGGGGCGTGGACCGGGAACACAGCGCTGCTGCTTTGTACAAGG 120

QY 157 GCTGAGTCATGCGTGGTGGTGGACATAGTTGCGAAGACAGTTCGCGCAAGCA 216

DB 121 GCTGAGTCATGCGTGGTGGTGGACATAGTTGCGAAGACAGTTCGCGCAAGCA 180

QY 217 GAACGAGCATGAACAGCAGTCCGAAGACGGAATCTCGGCTCCTACCGCGATTGGA 276

DB 181 GAACGAGCATGAACAGCAGTCCGAAGACGGAATCTCGGCTCCTACCGCGATTGGA 240

QY 277 AGACGAGACCGTCTCTGAGATGCGTTTGTGATGTCCTGTTCCGCGAGGATCTCGAACG 336

DB 241 AGACGAGACCGTCTCTGAGATGCGTTTGTGATGTCCTGTTCCGCGAGGATCTCGAACG 300

QY 337 GCGTGTGGGTGCTCATGCGTTCGTGATGCGAGTGGCGCAAAAGTCTCTGAACAGAGC 396

DB 301 GCGTGTGGGTGCTCATGCGTTCGTGATGCGAGTGGCGCAAAAGTCTCTGAACAGAGC 360

QY 397 TCCCGCGAGGCGAGTCAGAGTGGAAATGCTCTGTTGAAACCGTCATTTCGGTTTTTA 456

DB 361 TCCCGCGAGGCGAGTCAGAGTGGAAATGCTCTGTTGAAACCGTCATTTCGGTTTTTA 420

QY 457 CATTGTTTCGGGGCTATGATGACATGCCGGGCTTGTGCGTCCCGCTCAGCGACCGGC 516

DB 421 CATTGTTTCGGGGCTATGATGACATGCCGGGCTTGTGCGTCCCGCTCAGCGACCGGC 480

QY 517 CCGAACCACGAGATTCATGATAGCGCGGAACCCCTTAAGTCTCTTCTCGCAATG 576

DB 1177 CGAAGTGGATTGAGGAAAAACGCGACCTTCCCGAAGGAATGGTTGATCGCATACCCCGA 1200

DB 1141 CGAAGTGGATTGAGGAAAAACGCGACCTTCCCGAAGGAATGGTTGATCGCATACCCCGA 1200

QY 1237 CCGTTTCGCGGGAATCGCAAGAGCTCAACCGCGCGAGTGGCTGGATGACGACCTGC 1296

DB 1201 CCGTTTCGCGGGAATCGCAAGAGCTCAACCGCGCGAGTGGCTGGATGACGACCTGC 1260

QY 1297 CCGTTTCGCGGGAATCGCAAGAGCTCAACCGCGCGAGTGGCTGGATGACGACCTGC 1356

DB 1261 CCGTTTCGCGGGAATCGCAAGAGCTCAACCGCGCGAGTGGCTGGATGACGACCTGC 1320

QY 1357 CGCGCTTTGAAAAAGCGCGGTGCGAGATGCTGCTTCCCGAGGATTCGTGGTGGCT 1416

DB 1321 CGCGCTTTGAAAAAGCGCGGTGCGAGATGCTGCTTCCCGAGGATTCGTGGTGGCT 1380

QY 1417 AGATCCGAATGCTCAATGCGAGGCTGCTGCTTCCCGAGGATTCGTGGTGGCT 1476

DB 1381 AGATCCGAATGCTCAATGCGAGGCTGCTGCTTCCCGAGGATTCGTGGTGGCT 1440

QY 1477 ATGAGATGTTGATGAGCGCCATTGAAGACAGCAACCTCTTGGCAATCTGAAGAACTATC 1536

DB 1441 ATGAGATGTTGATGAGCGCCATTGAAGACAGCAACCTCTTGGCAATCTGAAGAACTATC 1500

QY 1537 TCACACAGGATGTCATCCCGACCTGAAAGCGCTTCAGGATGATGCGTCAAGAGCTATC 1596

DB 1501 TCACACAGGATGTCATCCCGACCTGAAAGCGCTTCAGGATGATGCGTCAAGAGCTATC 1560

QY 1597 GGGACAGCGTCATCAGCGCTTCTCCAAACAGGCGATGTCGGACACAGCGCTCGGATTG 1656

DB 1561 GGGACAGCGTCATCAGCGCTTCTCCAAACAGGCGATGTCGGACACAGCGCTCGGATTG 1620

1657 CTAGCGATGGCTGTTTCCAAAGTTTCAGGTGTTCTTGACGAGAAACCGTGGTCTGGGCGATCG 1716
1621 CTAGCGATGGCTGTTTCCAAAGTTTCAGGTGTTCTTGACGAGAAACCGTGGTCTGGGCGATCG 1680
1717 AAGACAAGCGGACCTGTACGTATAGCGTTTCGGAATTCATCTCAAAATGCTGC 1776
1681 AAGACAAGCGGACCTGTACGTATAGCGTTTCGGAATTCATCTCAAAATGCTGC 1740
1777 GTGTGTCGACGAGAAAGCGGGAGCTATGAATCGTCCGAGCCGACTTATGGCGACGCGC 1836
1741 GTGTGTCGACGAGAAAGCGGGAGCTATGAATCGTCCGAGCCGACTTATGGCGACGCGC 1800
1837 AATGGAAGTTGGCCAAAGCGGACGACTTCGAAAGCTCTCTCAAGCTCCCGCGCTTCGATG 1896
1801 AATGGAAGTTGGCCAAAGCGGACGACTTCGAAAGCTCTCTCAAGCTCCCGCGCTTCGATG 1860
1897 GGTGGCGGATCTGGATACGTTCCGACGTTGATCGATCAAAAGGTATGCTGCTGGGAGATCA 1956
1861 GGTGGCGGATCTGGATACGTTCCGACGTTGATCGATCAAAAGGTATGCTGCTGGGAGATCA 1920
1957 TCCGGAAGGCGGCTAAAGCGGCAATCCCGGCTGAATTCGGCTTTAGGGTAGCGAC 2016
1921 TCCGGAAGGCGGCTAAAGCGGCAATCCCGGCTGAATTCGGCTTTAGGGTAGCGAC 1980
2017 TGAACAGAAACCGCGCTCTGGAAGGAGCGCGTTCGCTTATGCTCAGATCTGTCCCA 2076
1981 TGAACAGAAACCGCGCTCTGGAAGGAGCGCGTTCGCTTATGCTCAGATCTGTCCCA 2040
2077 TCAGGACAAGGATCACGACGACCGACGATCAGGACAAGTCCGCTGGAGGGGAGCCCCAT 2136
2041 TCAGGACAAGGATCACGACGACCGACGATCAGGACAAGTCCGCTGGAGGGGAGCCCCAT 2100
2137 TCGAACGTAGCGGCATGACGCGCAGCGACCGAGATCAGGATTACAAGAAAGATCAGTCC 2196
2101 TCGAACGTAGCGGCATGACGCGCAGCGACCGAGATCAGGATTACAAGAAAGATCAGTCC 2160
2197 CATGGACATCTCTCTTCGCGGTTGAGACTGGTCTGTGTTCCGGGTGCTCAAAAGTTTC 2256
2161 CATGGACATCTCTCTTCGCGGTTGAGACTGGTCTGTGTTCCGGGTGCTCAAAAGTTTC 2220
2257 CGTAGGGCGGAAAGATCAAGCTGTCGGTTCGGCTTAATCCGTCGCCAAGCGCATG 2316
2221 CGTAGGGCGGAAAGATCAAGCTGTCGGTTCGGCTTAATCCGTCGCCAAGCGCATG 2280
2317 ATGCGGGCCACCGCTCTGTGCGCGTTTGGCTCTGTCTGACATAGGTTTCTGGGCG 2376
2281 ATGCGGGCCACCGCTCTGTGCGCGTTTGGCTCTGTCTGACATAGGTTTCTGGGCG 2340
2377 AGCAGTCCGGATGATGTTCCGGATCAGGTGCGGATGCGGACGCGGATTTCTGTGTCA 2436
2341 AGCAGTCCGGATGATGTTCCGGATCAGGTGCGGATGCGGACGCGGATTTCTGTGTCA 2400
2437 GTTGGCTGCGGTGATGCCAGAAATACGATAGGATCCGCTGTTCCGCTGGCGCG 2496
2401 GTTGGCTGCGGTGATGCCAGAAATACGATAGGATCCGCTGTTCCGCTGGCGCG 2460
2497 CGATTGTTGCGCTTTTCGGCCGCTCCCATGCTCTGCGGCGAGCCAAATGCCCCGTGA 2556
2461 CGATTGTTGCGCTTTTCGGCCGCTCCCATGCTCTGCGGCGAGCCAAATGCCCCGTGA 2520
2557 ACGGCTCGCAAAATCATGTTTCTTCGGGTGAAGCTGCGGCTGGGCGCGGATCGGCA 2616
2521 ACGGCTCGCAAAATCATGTTTCTTCGGGTGAAGCTGCGGCTGGGCGCGGATCGGCA 2580
2617 CGGGCGATACGAAACAGTCCGCTCATGAGTTTCTCAAGCGCGCGGATTTATCGGCATAG 2676
2581 CGGGCGATACGAAACAGTCCGCTCATGAGTTTCTCAAGCGCGCGGATTTATCGGCATAG 2640
2677 GCCTTGCCTATTCGCGGGCATACATCTCGAAATCGTCCGCTGGGCGCGGATCG 2736
2641 GCCTTGCCTATTCGCGGGCATACATCTCGAAATCGTCCGCTGGGCGCGGATCG 2700

2737 AACAGCATGCCGACTCTCTTGGTGTATCTGGGGGGAACACTGAAAGCAGGTCTTGAAGCG 2796
2701 AACAGCATGCCGACTCTCTTGGTGTATCTGGGGGGAACACTGAAAGCAGGTCTTGAAGCG 2760
2797 TTGATTTTCGTTCTCGTTTACCGGCGCGCTCGATCTTCGCCAGCTTTCGCCACAGGGCAACA 2856
2761 TTGATTTTCGTTCTCGTTTACCGGCGCGCTCGATCTTCGCCAGCTTTCGCCACAGGGCAACA 2820
2857 AGGCGGATGGCGTAAAGCTGATCTCTGTTGCCAGGCGCGCAAACTCTTGGCAGCGCG 2916
2821 AGGCGGATGGCGTAAAGCTGATCTCTGTTGCCAGGCGCGCAAACTCTTGGCAGCGCG 2880
2917 AAAAAGCGCGCTGTTGGGATCGGAGCGGCAATTCGCGGGAAGCGCTCACTCCAGCG 2976
2881 AAAAAGCGCGCTGTTGGGATCGGAGCGGCAATTCGCGGGAAGCGCTCACTCCAGCG 2940
2977 CCGCTTGAAGGCTTGAAGTAGCGAACCGTTATCGGGCGCATCCCGCAGGCTTGGCGCCATC 3036
2941 CCGCTTGAAGGCTTGAAGTAGCGAACCGTTATCGGGCGCATCCCGCAGGCTTGGCGCCATC 3000
3037 AGTGTCTCGAAAGGACCAACCAACCGCAAGCGCGACACCCAGCAACATCTTGCCTCAG 3096
3001 AGTGTCTCGAAAGGACCAACCAACCGCAAGCGCGACACCCAGCAACATCTTGCCTCAG 3060
3097 ATAGCCATGTCAATCAACCTAGCACGCCGCTCAACAGCGGCAAAATGACAGATCGCAGGCTA 3156
3061 ATAGCCATGTCAATCAACCTAGCACGCCGCTCAACAGCGGCAAAATGACAGATCGCAGGCTA 3120
3157 GGTGTAGGTGCTGATGCGCCAAACCGCGGCTTTCGGGTGTTGAGAGCTTAGAGTTAC 3216
3121 GGTGTAGGTGCTGATGCGCCAAACCGCGGCTTTCGGGTGTTGAGAGCTTAGAGTTAC 3180
3217 GAACCTATTCGCTCTCATGCTTTTGAAGCGCAGGCTTCTCTGTTCTGTTTTCATCAGCGAT 3276
3181 GAACCTATTCGCTCTCATGCTTTTGAAGCGCAGGCTTCTCTGTTCTGTTTTCATCAGCGAT 3240
3277 ATTTTATGCCCCCTTGATCCAGACTGCTACTTTCGATCCCTTTCGCTCTGATGAGAA 3336
3241 ATTTTATGCCCCCTTGATCCAGACTGCTACTTTCGATCCCTTTCGCTCTGATGAGAA 3300
3337 CTGATGATCTTTTGAATCAAGCGTCTGCCATGTTGGCTGACAGAAAGTCTCAACTGTTG 3396
3301 CTGATGATCTTTTGAATCAAGCGTCTGCCATGTTGGCTGACAGAAAGTCTCAACTGTTG 3360
3397 CGGAAGCGGATCATAAATGGGTTTCGATTCGCGGCGGCTGCTGTTTCATCTGGGCGC 3456
3361 CGGAAGCGGATCATAAATGGGTTTCGATTCGCGGCGGCTGCTGTTTCATCTGGGCGC 3420
3457 GTTCTGCTCATCTGCTGCTGTTCTGGGCTGTTGGATGCTGCGGCTGGGCTGATGTTGTT 3516
3421 GTTCTGCTCATCTGCTGCTGTTCTGGGCTGTTGGATGCTGCGGCTGGGCTGATGTTGTT 3480
3517 GCGCAGGATATTCGCTTCTCCGTCGCTTTCAGGGCGGCTTTCGGCTGGATCGAAGCT 3576
3481 GCGCAGGATATTCGCTTCTCCGTCGCTTTCAGGGCGGCTTTCGGCTGGATCGAAGCT 3540
3577 CAACATCGGATTCGCTGGGCTTTCGCGCAAAAGCGGCAAGCTAACCTTTCCTCTGG 3636
3541 CAACATCGGATTCGCTGGGCTTTCGCGCAAAAGCGGCAAGCTAACCTTTCCTCTGG 3600
3637 ACGTGTTCCTGAAGATGTTGTCAGTGTGCTGCAACCGCAGGGCTGAAGCGAGTGGCGCTCT 3696
3601 ACGTGTTCCTGAAGATGTTGTCAGTGTGCTGCAACCGCAGGGCTGAAGCGAGTGGCGCTCT 3660
3697 GGTGCTCGCGGATCGAGAGAACCCAGACAGCAAGCTCTGCTGCGGACTGCG 3756
3661 GGTGCTCGCGGATCGAGAGAACCCAGACAGCAAGCTCTGCTGCGGACTGCG 3720
3757 GCCATCGCTCCAGTATAGCCAGAACTCGGGTTCAGTGTCCAGTGTCCGAGCTCCGCTGCT 3816
3721 GCCATCGCTCCAGTATAGCCAGAACTCGGGTTCAGTGTCCAGTGTCCGAGCTCCGCTGCT 3780
3817 GACAGACAGGCTGCGCTTTCGAGGATCTACTCATTCGCGTTCGTTTCTCAAGCGCTTCA 3876

||||| 3781 GACAGACAGCGTCCGTTTACGAGATCACTCATTCGGTGTGTTCTCAAGGCGCTTCA 3840
QY 3877 AAGCCATTGTCGGTTTCGAAACATCAGGTCGGATCACTCAGCAGCTCCCGCGCAG 3936
Db 3841 AAGCCATTGTCGGTTTCGAAACATCAGGTCGGATCACTCAGCAGCTCCCGCGCAG 3900
QY 3937 AAGATATAAGGACGATCGCGGTCGGATGTCGGATCGGATCAGGACGATTAAGTACGAAC 3996
Db 3901 AAGATATAAGGACGATCGCGGTCGGATGTCGGATCGGATCAGGACGATTAAGTACGAAC 3960
QY 3997 CGGTTGCTCCATCCGTTTACCGGAGAGCCAGAAAAAAGCTCCGGAATGTCGCATTA 4056
Db 3961 CGGTTGCTCCATCCGTTTACCGGAGAGCCAGAAAAAAGCTCCGGAATGTCGCATTA 4020
QY 4057 TCCAGCCGACAGTCTCGAGTTTGGTCAATCAGCTCCGCGCGGCGCTG 4109
Db 4021 TCCAGCCGACAGTCTCGAGTTTGGTCAATCAGCTCCGCGCGGCGCTG 4073

RESULT 2
AB063188
LOCUS AB063188 1458 bp DNA linear BCT 19-JAN-2002
DEFINITION Gluconobacter oxydans sr gene for L-sorbose reductase, complete cds.
ACCESSION AB063188
VERSION AB063188.1 GI:18250946
KEYWORDS
SOURCE Gluconobacter oxydans (strain:IFO3291) DNA, clone:8.0 kb EcoRV fragment.
ORGANISM Gluconobacter oxydans
Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
Gluconobacter.
1
Shinoh,M., Tazoe,M. and Hoshino,T.
NADPH-dependent L-sorbose reductase is responsible for L-sorbose assimilation in luconobacter suboxydans IFO 3291
J. Bacteriol. 184 (3), 861-863 (2002)
21650687
REFERENCE 2 (bases 1 to 1458)
AUTHORS Shinoh,M.
DIRECT SUBMISSION
Submitted (13-JUN-2001) Masako Shinjoh, Nippon Roche Research Center, Applied Microbiology; Kajiwara 200, Kamakura, Kanagawa 247-8530, Japan (E-mail:masako.shinjo@roche.com, Tel:81-467-47-2226, Fax:81-467-45-6812)
FEATURES
Location/Qualifiers
1..1458
/organism="Gluconobacter oxydans"
/strain="IFO3291"
/db_xref="taxon:442"
/clone="8.0 kb EcoRV fragment"
/note="originally submitted as 'Gluconobacter suboxydans' (synonym: Gluconobacter oxydans subsp. suboxydans)";
1..1458
/gene="sr"
1..1458
/gene="sr"
/function="Reduction of L-sorbose to produce D-sorbitol"
/standard_name="NADPH-dependent L-sorbose reductase"
/note="synonym of NADP-b-sorbitol dehydrogenase (NADP-SLDH), homolog of NADP-SLDH with the accession number of AB028937"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="L-sorbose reductase"
/protein_id="BAB83933.1"
/db_xref="GI:18250947"
/translation="MITHETLKLSPAGVAPPYDINGIKPGIVHFGVGNFRFAHEAFY
VEQILKDPNGIIGVLTGSDRSKKAEFKKQDCLFSLTETAPSKGSTVRVMGALR
DYLLAPADPEAVLHKLADPGRIIVSWITEGGINNETTGFEDLENKAVQDDLTPEP
PSTIFGVVEGLRRRRONGRAFTIMSCDNLNRHNGNARKAFLGLYAKARDPELAKWIE

ENATFPNGMVDRIPTVSADIAKKLNEASGLHDDLPLVAEDFHOWLEDSFADGRPAL
EKAGVQGVGVDTYEHVKIRMLNAGHIMLCFPAVLAFENVDHSLADQRLRLLENFL
NKQIVPTLKKAPPGMTLEGYDSVTSRESNPAMADQTLRIISDGSXKIQVETVRKA
PEKRDLSRLTAFGMSYLEMRKDGKTGTYEPETFGDNHKTAKADDFESALKLP
AFDRWDETSGLNNKRVKRIKREKVAALPA"
BASE COUNT 329 a 386 c 434 g 309 t
ORIGIN
Query Match 23.2%; Score 955.6; DB 1; Length 1458;
Best Local Similarity 78.5%; Pred. No. 6e-183;
Matches 1144; Conservative 0; Mismatches 314; Indels 0; Gaps 0;
QY 537 ATGATTACGGCGAAACCCCTTAAGTCTCTCTCCCAATGTCAGGTCCTCCCTATGAC 596
Db 1 ATGATCAGCGACGAAACCCCTCAAGTCTCTCTCCCGCGGTGTGACAGGTCTCCGCTATGAC 60
QY 597 ATCAGCGATCAAGCCTGGATCGTCAATTTCCGTTAGGTAACTTTTTTCGAGCCCAT 656
Db 61 ATCAATGGGATCAAAACCGGGATCGTCAATTTTGGCGTGGGAACTTCTTCCGGGCGCAT 120
QY 657 GAGCGTTCTACGTGACGACAGATTTTGAACACACCTCCGGACTCGGCGATTTGTTGGTGT 716
Db 121 GAGCTTTCTACGTGACAGATCTCTCAAGACCATCCGAACCTCGGNAATCATCGCGTT 180
QY 717 GGCCTGACGGCAGTACCGCTTCAAGAAAAAAGCCGAGGAATTCAGAGCCAGGACTGC 776
Db 181 GGTCTGACGGGTACGACAGGTCAAGAAAGAGCCGAGGAATTCAGAAAGCAGGACTGC 240
QY 777 CTGTATTCCCTGACGAGACGGCTCGTCCGGGAAGAGCAGCGTGGCGGTCTATGGCGCG 836
Db 241 CTCTTTTCCCTGACGAAACCGCTCCGTCGGCAAGAGCAGCGTTCGTGTATGGCGCG 300
QY 837 CTGCTGACTATCTGCTTGCCCGCGCATCCGGAAGCCGTGCTGAAGCATCTTTGTTGAT 896
Db 301 CTGAGGATTAACCTTTTGGCTCTCTCGCATCCGGAAGCCGTGCTGAAGCATCTCGTGAC 360
QY 897 CCGGCCATCCCATCGTTTCCATCAGCATCACGGAAGCGCGCTACAAATCAACAGCAGACG 956
Db 361 CCGGAAATCCGATCGTTTCCATGACAATCACGGAAGCGCGTTTACAACATTAACAGCAGC 420
QY 957 ACCGCTGCTTCGATCTGGAGAATCGGCACATAAGGCCGACCTCAAGAACCCGGAAG 1016
Db 421 ACAGGTGAGTTTCGATCTTGAGAACAAAGCGGTTTCAGCAGGATCTGAGACACCCGAAACG 480
QY 1017 CCGTCTACCGTTTTCGGTTTACGTGTCGAGCGCCCTGCTGCTGCTTGGGATGCGCGTGT 1076
Db 481 CCGTCCACAACTTTTGGATATGTTGTGAAGACTGCGCCCGCGCTGACGACGAGTGGC 540
QY 1077 AAGCATTTACGGTCACTGCTGTGATAACCTGCGTCATACCGCAATGTGCGCCGCAAG 1136
Db 541 AAGGCTTTCAGCATCATCTCTCGGATAAATCTCGGCATACAGGTAATGTGCGCCGCAAG 600
QY 1137 GCCTTCTCGCTATCGAAGCGCGCATCCGAGTTGGCGAAGTGGATTGAGAGAAAC 1196
Db 601 GCATTTCTGGGATACGCGAAGCGCGTATCCGGAACCTGGCCAAAGTGGATTGAAGAGAC 660
QY 1197 GCGACCTTCCCGAAAGGAATGTTGATCGCATCACCCGACCGCTTTCGGCGGAATCGCC 1256
Db 661 GCGAGTTTCCCAATAGCATGTTGATCGCATCACCGACCGCTTTCGTGACATTCGG 720
QY 1257 AAGAGGTCAACGCGGCGCAGTGGCTGATGACGACCTGCGCGTGGTGGCGGAGATTC 1316
Db 721 AAGAGGTCAACGAAAGCAGTGGCTGCACGACGACCTGCGCGCTGTTGCAAGAGACTTT 780
QY 1317 CATCAGTGGTCTGGAGACGACGTTTCGGATCGCGCTCCGCGCTTGAAGAGCGCGC 1376
Db 781 CATCAGTGGTCTGGAGACGACGTTTCTGATGGCGGCTCGCGCTGGAAAGCGCGCA 840
QY 1377 GTGCAGATGGTTCGGGGAGCTGACGACTGGGAGTACGTCAAGATCCGAATCTCTCAATGCA 1436
Db 841 GTGCAGTCTGTTGGGATGTGACGACTACGAGCATGTAAATCCGATCTCTGAATGCT 900
QY 1437 GGGCATGTCTGCTCTGCCCGAGGCAATTCGTGCTCGCTATGAGATGTGGATGACGCC 1496


```
Db      1261 GGCACCTAGAGCATTCGAGCCGACGCTTTTGTGTGACAAACCATAGACTCTGGCCCAAGGCT 1320
QY      1857 GACGACTTCAAGCTCTCTGAAGCTCCCGGCTTCGATGGGTGGCGGATCTGGATACG 1916
Db      1321 GATGATTTTGGAGCGCGCTCAAGCTCGCAGCGTTCGATGCTCGCGGATCTGGAGACG 1380
QY      1917 TCCGAACTGGATCAAAAGTCTATCGTGTGCGGAGATCATCCGCGAAAGGCGGTAA 1976
Db      1381 TCCGGGCTGAACAAAGGTTCTGGAGCTTCGCAAGATTATCCGCGAAGGCGGTCAAG 1440
QY      1977 GCGGCATCCCGGCTCTGA 1994
Db      1441 GCTGCCCTCCGCGCTGA 1458

RESULT 5
LOCUS   AE004560
DEFINITION Pseudomonas aeruginosa PA01, section 221 of 529 of the complete genome.
VERSION AE004560 AE004091
KEYWORDS
SOURCE  AE004560.1 GI:9948372
ORGANISM Pseudomonas aeruginosa.
         Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
         Pseudomonas.
REFERENCE
AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
         Hickey,M.J., Brickman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
         Garber,R.L., Goltry,L., Tolentino,E., Westbrock-Wadman,S., Yuan,Y.,
         Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
         Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
         Complete genome sequence of Pseudomonas aeruginosa PA01, an
         opportunistic pathogen
JOURNAL Nature 406 (6799), 959-964 (2000)
MEDLINE 20437337
PUBMED 10984043
REFERENCE
AUTHORS 2 (bases 1 to 10977)
         Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
         Hickey,M.J., Brickman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
         Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E.,
         Westbrock-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
         Folger,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
         Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
         Hancock,R.E.W., Lory,S. and Olson,M.V.
         Direct Submission
TITLE Submitted (16-MAY-2000) Department of Medicine and Genetics,
JOURNAL University of Washington Genome Center, University Of Washington,
         Box 352145, Seattle, WA 98195, USA
FEATURES
source Location/Qualifiers
         1..10977
         /organism="Pseudomonas aeruginosa"
         /strain="PA01"
         /db_xref="taxon:287"
         /db_xref="taxon:287"
         107..1012
         /gene="mtlr"
         /note="PA2337"
         107..1012
         /gene="mtlr"
         /codon_start=1
         /transl_table=11
         /product="transcriptional regulator mtlR"
         /protein_id="AAG05725.1"
         /db_xref="GI:9948373"
         107..1012
         /translation="MPRAKVTDPSSYELMDHGHSLIYRHOHPPSLVRHFIKEYE
         LHLIVASSGKVIIGYICGNFAPDTLFLGPNLPHNWSIQVEPDEVVARDMLVNPTE
         VLEDSGVSENLAPLARYGIEFRPALGIESRRLLQRIADSGGMTRLGYFFI
         LMEQALCEDQLLSTVSSQLADEHNVERINRAVDYTFQHYAQLDTEQVAAHLGWT
         PTYFRFKAAGRGCFEIVNRLIRSKSCCELLARSELVPTVEVCFESGFSNLSNRRRE
         QOLKGMPPTGYRRLVTRLTQNL"
         1170..2480
         /gene="PA2338"

gene
source
         1170..2480
         /gene="PA2338"
         /codon_start=1
         /transl_table=11
         /product="mannitol dehydrogenase"

CDS
source
         1170..2480
         /gene="PA2338"
         /codon_start=1
         /transl_table=11
         /product="probable binding protein component of ABC
         maltose/mannitol transporter"
         /protein_id="AAG05726.1"
         /db_xref="GI:9948374"
         2556..3488
         /translation="MNDSIKALAAACIALPLAQAGFTLTITATVNNNDIMRMRLSK
         VFESHPIALKWVLEENVRORLTDTIATGGQFDLLTIGMYEAAKAGKAGLEPM
         SGLPADYALDDLPSVRGCLSVKGTLYALPYAERASITYYRKDLFQQAGLRPEQPTW
         TOLGEFAARLNRPQGYIGICLRKAGNGENMLIGTIANAFGARFDERWQPEFSGG
         EWKALDYYVSTLRKYPGPGSSNGENLALFNISGKCAIWDASVAGSFYDKSQSK
         VATGFAFAPREVTDKGASWLYSMALAIIPASSRAKDAKAFATWATSOAYGKLVADR
         EVANVPPTTRASTYSEAYLAAAPFARVLTLSLRKVDNPHPTLKPVPVGIQLVTYPE
         FOAICTOVGKLFSAALTCQMSDQALAAQAQSTAREMKRAGYK"
         2556..3488
         /gene="PA2339"
         2556..3488
         /gene="PA2339"
         /codon_start=1
         /transl_table=11
         /product="probable binding protein-dependent
         maltose/mannitol transport protein"
         /protein_id="AAG05727.1"
         /db_xref="GI:9948375"
         3499..4332
         /translation="MNTSALDTPLPARPARRLQRRRRPGMILLVSPSVALLLMMIV
         PLGWTLYPSLIYRNLLYPCENAFVGLNFTFVTVDAGFLPGALNTLLLVGSLVLSVA
         FGVLIAALKEAGFEFGAGRVLLISPFITMPSALLWKNLFIHPVSGILAAVMVRLF
         GAEPVDMUHAHYPLLSIILIVSNQWMLPFAILMTAMSLDQEQKEAARDGAGPLAIF
         WHLPLHARPIAVMIETIFLLSVFAEITFTTSGGPGYESTNLAYLIYNQALLQFD
         VGMASAGGLIAVLIANIAAIVLVRMIGNKLTERS"
         3499..4332
         /gene="PA2340"
         3499..4332
         /gene="PA2340"
         /codon_start=1
         /transl_table=11
         /product="probable binding protein-dependent
         maltose/mannitol transport protein"
         /protein_id="AAG05728.1"
         /db_xref="GI:9948376"
         4372..5484
         /translation="MLTKQSRRLRNALLGLLWGLALLFFPIFMMLLTFSKSEIDA
         FATPQFIFQPTLENYLHINERSDYFAYANSLLSISSATLLCLMVAVPAAYSAFFE
         TRHKRTLLMWLSTKMLPPVGLMPYLLAKQFGLLDSRLALITVYTLINLPVVMVW
         YTYFKDIPVDILEAARLDGACTWQEIIVRVLPIARGGLASTLLLSLILCWNEAFSLN
         LTFSSAAPTALVASYSPEGLFWAKLSAVSTLACAPILVFGWISQKOLVRGLSFGAV
         K"
         4372..5484
         /gene="PA2341"
         4372..5484
         /gene="PA2341"
         /codon_start=1
         /transl_table=11
         /product="probable ATP-binding component of ABC
         maltose/mannitol transporter"
         /protein_id="AAG05729.1"
         /db_xref="GI:9948377"
         5507..6982
         /translation="MADLKIRNLOKFGDQALIKGIDLDVRDREVFVVGPSGCGKST
         LRLIAGLEASGGSIALDGDITDTPPAKRDLAMVFQTYALYPHMTVRNLSFALDL
         AGVDKREVAARVADAAARTELOALIERKPROLSGQORVAIGRAIVRNPKIFLFDPE
         LSNLDAALRVQMRLEARLHQELATMIYVTHQVAMTADKVVVLNPRGVEQVGP
         LELYHPANLFVAGFLGTPKMGFLGRHLSRNGERCEVALECGARVGLPTLCAGELATG
         SPVTLGRPEHLNIRARPCAFQGLLRVAADVSRILGSDSFCHVRADSGEMLTLVRVGD
         PAPDYGRAPLELGFDPDEHCHLFDNSQALDKRLQQA"
         5507..6982
         /gene="mtlD"
         5507..6982
         /note="PA2342"
         /gene="mtlD"
         /codon_start=1
         /transl_table=11
         /product="mannitol dehydrogenase"
```

```
/protein_id="AAG05730.1"
/db_xref="GI:9948378"
/translation="MKLNROHLLPLATAVARPSYDPAQLRQGIHVHIGVGGFHRAHQAA
YTDALMRGALDAWICGALSRSDRAMHDALAAQDLYTLIELGDDPDTEVRVIGAI
SMLLAEDGGALEALEKLAEPARIVLSITITEGVCIDGSGEFLAELPLVRHDLNPR
TPRGVFLCEALRRRDRDTPAFVTVMSDNLPHNGVARKALLAFERLDPLGLRWI
ATHVFNAMVDRIITPMKISPAHRHDLQAORHDVEDAMPVCEPFVQWVLEDRFSAGRPA
WEKVGQFTDDVTPEEMKIGLLNSHSLATLGLFRYFRVHETLGDPLLRVYRAF
MDRVPALLAPVPCIDLERKYKDSLVERFANRAIADOLERVCSDGSSKFPKFIPTANR
LIAAGRLERVALVVAWALYLGVDHGERYPIDPRAAEQALVAERHGLGRRLIG
VETVFGPAIPASAAVFAFERLDLSLRHGVSETLRQVGE"
/length=8487
/gene="mtly"
/notes="PA2343"
6979..8487
/gene="mtly"
/codon_start=1
/transl_table=11
/product="xylulose kinase"
/protein_id="AAG05731.1"
/db_xref="GI:9948379"
/translation="MTPEGRMIGLFLGICGTGSKALLLDAGSGRTLGLGSAORPP
EGRGRREOPADMLWEAMASAVRNALPEAAVDGREVALAVSAQOHLILLDAGRAL
PAKLLDTSNAENRELDALGGPAGSLERLGLVLPAGYTLKLLNSRRRFPLELFR
VAHLLPHDYLNHWTGRVCEAGDASGSGFYDVRRTWASDVLDELVEPGRLAAALP
ELIEPGACIGNLRPEAAALGAPHTRVAGGGDNMLAAITGNIRPGLLTASLGTSG
TLSAYABRLVSPHGEIATFCASSGWLPLACTMNLTCAGLVODLLHLDLDEFSLA
AQAPEGGLMLPFFDGERVAPLPHASASLHGMTAANLSRANLCRAVLLEGTAFLRY
GLDLRASGLPGEIRLVGGAKNPLNRRTLADLLGLPLVCPRTEAAALGAALQAAN
SLGRESAGESLEALRCVALDESTRTQPARQAAEQAYRYLEHLP"
/length=9460
/gene="mtly"
/notes="PA2344"
8528..9460
/gene="mtly"
/codon_start=1
/transl_table=11
/product="fructokinase"
/protein_id="AAG05732.1"
/db_xref="GI:9948380"
/translation="MYLVGEALFDVFSLEARSNELGFTAIAGGSPFNVAVLRLR
GVEALFGGLSSDYLGLRLRVLEEGVDCGLFVPSDAPPTLAMVGLDASGAYQFR
GEGCARVLEHLPDLGRIRGLHVGSYTLVTPVADTLLALVRSGRLVSLDPN
VRLQDPDIDLRWRVEAFASHAHLIKASEEDLALLYPGRDPGEVARGMLPRCLVF
VTHGAGASVHCTGSRSPADTALPLRDTVAGDFTQAAATLAYLRILDADSPAGLAA
LSRAIDAMLAFAIRAAAVTCSRVPDLPFAHEL"
/complement(9503..10738)
/gene="PA2345"
complement(9503..10738)
/gene="PA2345"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG05733.1"
/db_xref="GI:9948381"
/translation="MQSANASELFTDIAVIGGGAAGTCVIAASLLKRSPLRISLIEPA
DTHYQPGTIVGGGAYQARTPMAGLVPPGEWLRTRVDRVDPPEARLLLEGGDS
LEYRLNLPCPLRLEWIEGLEITLGRNGVTSNYDLAPYTWELVRLGRLEKALFT
OPAMPKICAGAPKAMYLSCDHLREGVLQDIEVEFDLAGAALFGVADFPVPLMEYVR
KYSALFNSNLKVDGAARKAWFEVKDADNTNIAEKDFLLHVVVPOLPPTFAAS
GLGDAAGCEVDPATLQHVRIEYFALGDVCGTANAKTAAARQIIVVAENLLGLRS
GRLPRYDVGCGPLTVVERGRVLADEFYAGKLLPFPLEPTAASRFANLLKRLVLP
WYWNMLKLGREWLPGR"
/length=1704
/gene="mtly"
/notes="PA2346"
1674 a 4067 c 3532 g 1704 t
BASE COUNT 1674 a 4067 c 3532 g 1704 t
ORIGIN
Query Match 8.5%; Score 351.4; DB 1; Length 10977;
Best Local Similarity 55.5%; Pred. No. 8.4e-61;
Matches 726; Conservative 0; Mismatches 571; Indels 12; Gaps 2;
QY 564 CTTCTGCTCAAGTGTCCAGGCTCCCGCCCTATGACATCGACGGGATCAAGCCTGGGATCGTG 623
II | II | II | II | II | II | II | II | II | II | II | II | II | II | II |
Db 5537 CTGGGACCGGTGTGCGCCGCCCTCTCTACGACCCCGGCCCACTGGCCGAGGCGATCGTC 5596
```

QY 1698 ACCGTGCGTGGCGCATCGAAGACGCGGACCTGTCACTATAGCGTTCGGAATTGCA 1757
 Db 6671 ACCGCCAATCGGCTGCGCGCGCGCGCTGGAACGCGTGGTGGTGGCG 6730
 QY 1758 TCCTATCTCGAAATCGCTGGTGGCGCACGAGGCGGACGATGATGATCTCGGAG 1817
 Db 6731 CGCTGGGCGCTGTACCTGGCGGGTCCGACGATGCGCAACGCTACCCGATCCCGAT 6790
 QY 1818 CCGACTTATGCGCGCGCAATGGAAGTTGCCAAGCGCGGACTTCG 1866
 Db 6791 CCGCGCGCGCGCAATGCGCGCGCTGTGCGCAACGCCGCGCTCG 6839
 RESULT 6
 AF007800
 LOCUS
 DEFINITION
 Pseudomonas fluorescens mannitol operon, MtLE (mtLE), MtLF (mtLF),
 MtIG (mtIG), MtLK (mtLK), mannitol dehydrogenase (mtID), xylulose
 kinase (mtLY) and fructokinase (mtLZ) genes, complete cds.
 ACCESSION
 VERSION AF007800 U39468
 KEYWORDS
 SOURCE AF007800.1 GI:2293413
 ORGANISM
 Pseudomonas fluorescens.
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.
 1 (bases 1 to 10031)
 Brunker, P., Altenbuchner, J., Kulbe, K.D. and Mattes, R.
 Cloning, nucleotide sequence and expression of a mannitol
 dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in
 Escherichia coli
 Biochim. Biophys. Acta 1351 (1-2), 157-167 (1997)
 2 (bases 1 to 10031)
 Brunker, P., Altenbuchner, J. and Mattes, R.
 Structure and function of the genes involved in mannitol, arabinol
 and glucitol utilization from Pseudomonas fluorescens DSM50106
 Gene 206 (1), 117-126 (1998)
 9461423
 3 (bases 1 to 10031)
 Brunker, P., Altenbuchner, J. and Mattes, R.
 Direct Submission
 Submitted (10-JUN-1997) Institute of Industrial Genetics,
 University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany
 On Aug 4, 1997 this sequence version replaced gi:2065484.
 FEATURES
 Location/Qualifiers
 1..10031
 /organism="Pseudomonas fluorescens"
 /strain="DSM50106"
 /db_xref="taxon:294"
 814..842
 932..2242
 /gene="mtLE"
 /note="similar to E. coli periplasmic maltose binding
 protein MtLE; periplasmic subunit of the periplasmic
 binding protein dependent mannitol transport system"
 /codon_start=1
 /transl_table=11
 /product="MtLE"
 /protein_id="AAC04468.1"
 /db_xref="GI:2293414"
 /translation="MKFTAKALLACTCMTLSAVSLGAQTLTATVNSDMIRMOKLSK
 TFEAHPKILNWWVLENVLRQITTDIATGGQFVDLTIGMYEALWGAAGKLEPM
 KDLPSYDLDFVSVRDGLSVKSLXALPYAESSTIYRTDLFKDAGLSMPEHPTW
 SQIGFAAKLTDKSKQYGLCLRGKAGNENALITLANGYGARWFEKWPQEFNG
 EKDALNIFYVDMKMGKPPGASSNGENLALFNSGKCAIWDVASVAGSFVTDKTSK
 VADHVGTFEAPHEKTDKTSWLSLAIPTSSKAKDAKAVFTSWATSKYEGALVAKT
 DGVANVPCTKKSTYSDEYKMAAFPAKVLTLESKADPTKPLKPVIGIQLVTYPE
 FOAINTQVGFESGALTGGQQTVDAAALTAQAQTTRREMKGAGYK"

gene
 CDS

2381..3271
 /gene="mtLF"
 2381..3271
 /gene="mtLF"

/note="similar to E. coli MalF protein; inner membrane
 subunit of the periplasmic binding protein dependent
 mannitol transport system"

/codon_start=1
 /transl_table=11
 /product="MtLF"
 /protein_id="AAC04469.1"
 /db_xref="GI:2293415"

/translation="MNTSAKNRLANPGWFLVTPSVAALLLMVPLQMTLVFSIRYN
 LLYPENQFVGLNFTYFIDSGFLPGATNLLVSVLLISVYVGVLSALLEASEF
 LGRGIVRVLLISPFIMPTVGALIKWNLIFHPVSGILAAVWKFAGPVDWLAHPIL
 VVITISVOMLPEAIIILMTAMOSLDQEKAAERLDGAGATAFMHLLPLHARPIAV
 VMLETILSVFAEIEFTTNGGPGYASTNLAYLIYNOALVQFVDVGNASAGLLAVVI
 ANIAAILVRMIGRNLTKDP"

3282..4112
 /gene="mtIG"
 3282..4112
 /gene="mtIG"

/note="similar to E. coli MalG protein; inner membrane
 subunit of the periplasmic binding protein dependent
 mannitol transport system"

/codon_start=1
 /transl_table=11
 /product="MtIG"
 /protein_id="AAC04470.1"
 /db_xref="GI:2293416"

/translation="MTLOQSRNSRQSLLLGLTLAWAIALIFFPFMVLTSFKTEIDAF
 ATPQFIFTPTLENYLHINRSNYFSYAWNSVLISFSATALLISVPAAYSMAYET
 KRKSTLLMLSTKMLPPVGLMPLIYLAKSGGLDTRIALIITILNLPVVMVY
 TYPKDIPKIDLEAARLDGATLWQEMRVLLPIAKGLASTVLLSLILCWNFAFWSNL
 TSSNAAPLTALIASYSSPEGLFWAKLSAVSTLACAPILIFGWISQKQLVRLSFGAVK
 "

4136..5239
 /gene="mtLK"
 4136..5239
 /gene="mtLK"

/note="similar to E. coli MalK protein; ATP binding
 protein subunit of the periplasmic binding protein
 dependent mannitol transport system"

/codon_start=1
 /transl_table=11
 /product="MtLK"
 /protein_id="AAC04471.1"
 /db_xref="GI:2293417"

/translation="MANLKIKNLQKGFEGFSIIKIDLEVNDFVVFVPGSGCGKST
 LLRLIAGLEEVSEGTIELDGRDITEVTPAKRDLAMVFTQYALYPHMSVRKNMSFALDL
 AGVDKLVESKVNAAARILELGPLLRKPKLGGQQRVAIGRAIVRNPKIFLFDPEP
 LSNLDAALRVQMRLELARLHKELOATMYVTHDOVEAMTADKVVVLNSGRIEQVGP
 LELYHPANLFAVAGFLGTPKMGFLKGVTRVDGCGCEVQLDAGTLISLPSGLSVG
 SAVTGLIRPEHLEIASPGQTTLTVTADVGERLSDGTFCFCHVITSNGEPLTMKIRGDMAS
 QYGTLLHLDPAPACHLFDTDGVAVAVPLRAA"

5258..6739
 /gene="mtID"
 5258..6739
 /gene="mtID"

/EC_number="1.1.1.67"
 /note="MtID"
 /codon_start=1
 /transl_table=11
 /product="mannitol dehydrogenase"

/protein_id="AAC04472.1"
 /db_xref="GI:2293418"

/translation="MKLNQNLTLAPEVKLPAYTLADTROGIAHIGVGGFPHRAHQAY
 YTDALMTGEGLDWSICGVLRSRDKARDLADGQDYLFITYELGDDTDEVRVIGSI
 SDMLLADSQAALDKLASPEIRIVSLTITEGGYCIDDNGCFMAHLPOIHOHLAHP
 SPKTVFGICAALTQRAAGIPAFVMSCONLPNGAVTRKALLAFAALHNAELHDI
 KAHVSFFNANVDRITPMTSTAHLQLHDEHGIDDAWPVCEPFQVNWLEDFVNGRPA
 KHVGVGFTDDVTDEYEMKICLLNGSHLALTYLGLFKGYREVHTEHMDPLFVATMAY
 MLDLVTNLPAPVPGIDLTQYKQTLVDVRSNOAIDQLERVCSDGSSKFPKFTVPTINR

REFERENCE	Phyllobacteriaceae; Mesorhizobium.		
AUTHORS	1 (sites) Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idegawa, K., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kiyokawa, C., Kohara, M., Matsumoto, M., Matsuno, A., Mochizuki, Y., Nakayama, S., Nakazaki, N., Shimpo, S., Sugimoto, M., Takeuchi, C., Yamada, M. and Tabata, S.	gene	
TITLE	Complete genome structure of the nitrogen-fixing symbiotic bacterium Mesorhizobium loti	CDS	
JOURNAL	DNA Res. 7 (6), 331-338 (2000)		
MEDLINE	2 (bases 1 to 332635)		
REFERENCE	Kaneko, T.		
AUTHORS	Submitted Submission		
TITLE	Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research Institute, The First Laboratory for Plant Gene Research; Yana		
JOURNAL	1532-3, Kisarazu, Chiba 292-0812, Japan (E-mail: kaneko@kazusa.or.jp, rhizobase/, URL: http://www.kazusa.or.jp/rhizobase/, Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)		
COMMENT	On May 11, 2001 this sequence version replaced gi:11994980.		
FEATURES	Location/Qualifiers		
source	1..332635 /organism="Mesorhizobium loti" /strain="MAFF303099" /db_xref="taxon:381" complement(79..924) /gene="ml14780" complement(79..924) /gene="ml14780" /codon_start=1 /transl_table=11 /product="2-hydroxyhepta-2,4-diene-1,7-dioate isomerase" /protein_id="BAB51359.1" /db_xref="GI:14024757"	gene	
gene	/translation="MKLRGVEGSESRPGLLDADGTTIRDLISAHVADIAGTTLHPASLE MLGLDKRSPAVSGACPRGACVAGTKFCIGLNYSDHAAETGATVPPEPIPMKAS SAIIPDDDLIPRGSEKTDWEVLAVIGTKATKYVSEADALDYAGVAVAHDSERA FOAERQGWTKGKCDTPTGTPMLVTKDEVADPNLKMWLTVMKGTWQNGSTKTMVY GVKLVSVLSQFMSLHPGDIISTCTPPGVGLGMKPPVFLRAGDVVELGIEGLGQOKOT FKADV"	CDS	
gene	complement(1117..1848) /gene="ml14781"		
CDS	complement(1117..1848) /gene="ml14781" /codon_start=1 /transl_table=11 /product="oxidoreductase" /protein_id="BAB51360.1" /db_xref="GI:14024758"	gene	
gene	/translation="MADLAGKVVVITAAAGIGKASALAFAKTATGATVHATDINETLLA ELAKTPGKTRKLDVLDNAEVNTTFAEIGRVDVLFNCAGFVHSGSILEMKDGLDFAF NLNVRAMRTIRAVLPGMLRGDGSIVNMSVACAGKGVNPRFAYGVTKAIVGLTKA IAADYVKGKTECNACIGTGVESPSLDQMIAQGYEARAAFTIARQPMGLTPEETA DLAVLAGATTISQANIDGWSI"	CDS	
gene	complement(2038..4383) /gene="ml14783"		
CDS	complement(2038..4383) /gene="ml14783" /codon_start=1 /transl_table=11 /product="aldehyde dehydrogenase" /protein_id="BAB51361.1" /db_xref="GI:14024759"	gene	
gene	/translation="MEYGAPEARENEADWLAAEDFGKALFIDGAKKAAASGGKTFDTS EPSSGKLLAKVSDASADIDAVAAATKALPKWSASSGYARAKVLYIAIGRAMQRRL FVLESIDNKGPIRESRDIDVPLAIRHFIHAGWAQALDRDPCHKGKGVVQGIIPWN PFLMLAKKAPALAAQCTVYLKPAEFTPLTALIFAETCEAGVPKGVNIVQGPFA GAATVHPGQIKTAFTGSSEVGKIRKATAGSGKKLSLELGGSAFTVFEADLDSAV EGLVDGTFWQGVCCAGRLLVOEGIAEALIAKVKTRMSRLRVGSPDLKNTDGLPLV DLTOLDRVKGLVAGAKQGVACWOPDAALPSSGGYHLPTLATGVS PANILIAQEVFGP VLATMFRNTEEALELANRGTIVGLAASVSENVNLAHVAPQLKAGVWVWNGTNMFA ACGFGGYSRESGREGGMEFYLTAKPLGPVIRKPAAVSAQPEQADGAADRTAK	CDS	
gene	complement(5705..6532) /gene="ml14785"		
CDS	complement(5705..6532) /gene="ml14785" /codon_start=1 /transl_table=11 /product="NADPH:ferredoxin reductase" /protein_id="BAB51363.1" /db_xref="GI:14024761"	gene	
gene	/translation="MNTTSAPNTAGARPLOQPIIPANYAETVSVKHYTDLRFSFRIT RPSQLRFRSGEFVMIPLNPAEKPVFRAYSVASPAWDELEFFSIKVPDGPLTSELQKI QVGTDTVMROKSTGTLVVDALTPAKRLFMISTGTGFIAPFASLLRDPDTYKFDOLILT HCTDNELRYGOELVAALESDDLIGLTTGRVPLYNSTTRESARMGRITALLGSGK FYADLGIDKLNKPEITDRIMICSMHMLKDVNELAESLGFQEGSLSHPATFVVERAFVG" 6753..8837 /gene="ml14786"	CDS	
gene	complement(5705..6532) /gene="ml14786"		
CDS	complement(5705..6532) /gene="ml14786" /codon_start=1 /transl_table=11 /product="N-methylhydantoinase A; mikimopine-lactam hydrogenase" /protein_id="BAB51364.1" /db_xref="GI:14024762"	gene	
gene	/translation="MKNFVSPPDSLSGVAAGIDVGSTFTDLLLIDRGGKVIHAK TPTTVNQAGVVSALGATGFAIDGIDLIIVHGTTTTTNAVLERLARTGHTITGFRD VIELGRRTRPOPYGTMGTGTFVPIPRNLRLEVERSEASGAVRTPDDEAMEAEVATLI AAGCESLVIHFLHSYANPAHERAAEIAAELWPNGYITTHALLSEAREFEFVTSV NASQPIILERYVERLRKELAAKGYARDFLIMNGNGMISARFVRESAKTVMSGPASG VIAAAYTKKAGFENLVLYDMGGTSTOVALIRNAEPASVNEIEIETAYAMPIHVPAVH TVGAGGGSIAERVDAAGLIQIGPESAGANPGPICYGRGLLEPTITDANLVLRAPKLL LAVDNPVTVERVTCIFADKIGKATGLSCVEAGAVLRLGNMKAGAIRMYSVSRGHP RDFALFAFGAGPLHATALRGLGPRVLVPAREGITNALGCVVADLRHDFVNTINOP VAVDETLQHLVLEHRNEGEQLIGKAVKAPETIRVTHSDAMQVFGQTHIINVPSPS TVTRQELQALFEKAYFARFVLEPEIRANLVNLTSVTGRPAIDLRLIDPAGRAKT LDEARREIRPVYAGOWHDTVPYAREKPLDPLDAVIEGPAILEQMDATTVPLEPDRARSD ADGNLIIDIGEFA" 8854..10596 /gene="ml14788"	gene	
gene	complement(5705..6532) /gene="ml14788"		
CDS	complement(5705..6532) /gene="ml14788" /codon_start=1 /transl_table=11 /product="N-methylhydantoinase B" /protein_id="BAB51365.1" /db_xref="GI:14024763"	gene	
gene	/translation="MAKLDATISLVQAALQOQVCEMDLTFSSRAAFSPVIAEANDRS GIYSADVSLIAQSGGLPVFVGMVYSTKTVIEMADGRCLEPEPGDVIYNPDYLG GTHLMDYFAMPYVRACKIECNLSNTHGWDPDGGSPGGSATATAVEQGLRLPPYK LFRKGVLDPEIYAIACISNIRVADORIGDIRAQAALLIQDRLNGILIDYQDETFVVEA IAELRRRAAOMRANIAIPDGYTHSAFVDSGVNEPEPTIALAVSKQGDITLTFDA GSSKPCAPMNSVLATTLSSVYLAMRHIFPDVPIISAGAFPLIVRRPEGTFLDARYPR	CDS	


```

PVSCAAEVQSRIAEVFAAMVQALPKDVTAAAPAGSSGNFALGNDPARGRDYVWYQI
SGGGYNGSHDGTNGCTTIGSKSPPEIMEAQFVLYRHYALRSGSGGAGKHRGG
FGLAYEILRGDARAGFVMDHGRFGQALGKGDGPNSVTVPFRGVAVHPPLSLKE
QDILPKAGDRVRVCTPGGGYGDPRERDADQVQDVALGYTAEASEKFCGVLSADG
LTVRKATDGGORAG"
10720..11493
/gene="mlr4789"
10720..11493
/gene="mlr4789"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter, periplasmic amino
acid-binding protein"
/protein_id="BAB51366.1"
/db_xref="GI:14024764"
/translation="MRLFGREVLAAVALTLASGAASQDKLRIGTEGAYPPFNANA
DGTGGFDIDLGKALCEMKACEFTVQDFGCSIPALQAGKFDALINHTIPERAKEV
EFTHYOTPAITVPKDSITAGTSDLLKGLVQVATIHQFASOKYSGSVTKAY
PTGDARTDMANGRLDAMMDGSIILTEMLKPKDGACCKLLGLTADPAIHGPGVIAL
QKNGELAGFNAADALRANGKYKEINDKYFSDVYCN"
11674..12801
/gene="mlr4790"
11674..12801
/gene="mlr4790"
/codon_start=1
/transl_table=11
/product="aminomethyltransferase"
/protein_id="BAB51367.1"
/db_xref="GI:14024765"
/translation="METFLAERRRPFYSSIVGLGATMGRVGGDFISAKYSGVTDHLL
NTRVQVQDLSLTKMDIKGPDAAELVNHVINDAVAMKPGQVRYSTVCREDDGIND
DLTVRGLGPEHFMLVTGSVNRKMLPWLQHAQKAYVTDITAAVAPPTIQGPSRE
LLKAMIADADLDKRWAFTHGSHVNGTVRLTSRTGVTGELGFLVPADEAASVMDTL
MRAGDKFLKPYGLVAMFTLGLKAYPAHJDMDETPTPFHVGLDRWKPKDGFICR
EALLKIRKDGIDERTWTLGLLDGNKPAATDARVLADGEDAGIVTSDHCYSLGKVLATA
HLRPLFAIGTELSIDIGRPRETRAVVAPMPFFDEGVKLRG"
12955..13611
/gene="ml4791"
12955..13611
/gene="ml4791"
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="BAB51368.1"
/db_xref="GI:14024766"
/translation="MSPILFELLRSIWETVLTMAASGLISLFLPLCLALISTERG
7.1%; Score 292.4; DB 1; Length 332635;
Best Local Similarity 55.1%; Pred. No. 5e-49;
Matches 616; Conservative 0; Mismatches 496; Indels 6; Gaps 2;

QY 551 AACCTTAAGTCTCTCTCGCAATGTCCAGGCTCCGCCCTATGACATCGACGGATCAA 610
DB 107979 AAACCTCCAGCCTCCCGGCAATGTTCGCGTCCGGAATACGACCCCTCGAAGCTCAA 107920

QY 611 GCCTGGGATCGTGAATTTCCGTGTAGGTAACTTTTTCGAGCCCATGAGCGGTCTACGT 670
DB 107919 GCCTGGGATCGTGAATTTCCGTGTAGGTAACTTTTTCGAGCCCATGAGCGGTCTACGT 107860

QY 671 CGAGCAGATCTTGNACAGCTCCGGACTCGGCATGTTGTTGGTGGCCCTGACGGGCGAG 730
DB 107859 CGACGATCTGTTCAATTCAGGTGTGGGCCATGACTGGCGCGTGTGCGCGCAGCGGTGT 107800

QY 731 TCACCGTTTCAAGAAAAAGCCGAGAAATCAAGGCCCGCAGGACTCCCTCTATTCCCTCAC 790
DB 107799 CGAGGGCCAGAGATCGTTCGCGGCAAGCTCGAGAGCAGGACTGGCTGACCCCGTGT 107740

QY 791 CGAGACGGCTCCGTCGCGGCAAGACACGGTGCAGGTGATGGCGCGTGCCTGACTATCT 850
DB 107739 CGAGCAGAT---GAAGGCCATATGAGCGCCGCTGTCACCGCGCGGATGATCGATTTCT 107683

QY 851 GCTTGCCCGCGCGATCCGAGCGCGTCTGAGCATCTTGTGATCCGGCCATCCGCAT 910
DB 107683 GCTTGCCCGCGCGATCCGAGCGCGTCTGAGCATCTTGTGATCCGGCCATCCGCAT 910
```

```

Db 107682 GAT---CCCTGGCGATCGGCGCGCATCATCGACCGGTGTGGCCGACCGCGATCAAGAT 107626
QY 911 CGTTTCCATGACGATCAGGAAGCGCGCTCAACATCAACGAGACGACCGGTGCTTCGA 970
DB 107625 CGTTTCCATGACGATCAGGAAGCGCGCTCAACATCAACGAGACGACCGGTGCTTCGA 107566
QY 971 TCTGAGAAATGCGGACGTAAGGCGCGACTCAAGAACCCGGAAGACCGCTTACCGT 1030
DB 107565 CCCAGCCCATCGGATATCGTCGCGGACGCCAACCCGCGCGCCACCCCAAGACAGTGT 107506
QY 1031 CGTTTACGTGTCAGGCGCGCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1090
DB 107505 TGGCATCATCTGCGAGGACTGCTGCGCGCGGACGAGGACCGCTGCGGTTCACCGT 107446
QY 1091 CATGCTCTGTATACCTGCGTCAACGCGCAATGTCGCGCGCAAGGCGCTTCCTCGGCTA 1150
DB 107445 GATGCTCTGCGACCAATTCGCCACACGCGGCTGCTCACCTCGGATGGCTGCTGCTGCT 107386
QY 1151 TGGCAAGCGCGCGCTGCGGAGTGGCGAGTGGAGTGGAGAAACGCGACGCTTCCCGAA 1210
DB 107385 GCGCGCGTGCATGATGAGGATTTGGCGGTTGGTGGTGGAGCAGCAACGCTGCGCTTCCAAA 107326
QY 1211 CGGAATGTTGATCGCATCACCGCGCGCTTTCGCGGGAATCCGCCAAGAGCTCAACGC 1270
DB 107325 CGGCATGTTGAGCGCGCATACGCGCGGCAACACAGACCGCGGCGGCTACTGCGCAA 107266
QY 1271 GCGCAGTGGGCTGATGACGACCTGCGCGTGGTGGCGGAGGATTTCCATCAGTGGGTGCT 1330
DB 107265 GGAATTCGGCTCGAGGATGCTGCGCGGTTCTGCGAGCGCTTCCGGCAATGGGTGCT 107206
QY 1331 GGAAGACAGTTTCGGATGCGCGCTGCGCGCTTGAAGAACCGCGGCTGCGAGTGGTGG 1390
DB 107205 GGAAGACAGTTTCACCGATGCGCGCTGCGCGCTGGAAGAACCGCGGCTGCGAGTGGTGG 107146
QY 1391 GGAGCTGACGACCTGGGAGTACGTCAGATCCGAGTCTCAATCAGGCGCATGCTATCCT 1450
DB 107145 GGATGTCGCGCGCTTACGAGTGGATGAAGATCCGATCCCTCAATCAGGCGCATGCTATCCT 107086
QY 1451 CTGCTTCCAGGCAATTCGTCGCGCTATGAGAAATGTCGATCAGCCTTGAAGACAGCGA 1510
DB 107085 CGCCTATCCGCGCGGCTGATGAGATCCATTCCTTCATCAGGCGGATGAGGAGCGCT 107026
QY 1511 ACTCCTTGGCAATCTGAAGAACTATCTCAACAGAGTGTATCCCGACCTTGAAGGCGCC 1570
DB 107025 GGTGCGCGGCTTCTGACAAAGCTCGAGCAGCAGAGATCATCCGCGGCTGCGCGCGT 106966
QY 1571 TTCAGGATGACGCTCGAAGGCTATCGGAGCAGCTATCAGCGCTTCTTCCAAACAGGC 1630
DB 106965 GCCGACACGCTGCTGAGGATTTATACCAGCTCATCGAAGGCGCTTTTCCAAACCCCAA 106906
QY 1631 GATGTCGACACGCTCGCGATTCGCGGATTCGTCGAGGATGGCT 1668
DB 106905 GATGTCGACACCATCCGCGGCTCTGCTCGACGCT 106868

RESULT 8
SME591790/c 323450 bp DNA linear BCT 05-JUL-2002
LOCUS Sinorhizobium meliloti 1021 complete chromosome; segment 9/12.
DEFINITION AL591790 AL591688
ACCESSION AL591790.1 GI:15075230
VERSION
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 323450)
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelie,D., Puehler,A., Purnelle,B., Ramsperger,U.,
Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
TITLE Analysis of the chromosome sequence of the legume symbiont
```


JOURNAL	Sinorhizobium meliloti strain 1021	gene	KAHALSGEGFTGALLDRRAGTIQIOLAYARGLARAAALAAAGAEIETDTPLLAASQGGDL
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)	CDS	WNLTGPRGTVTARHVLATINAYGSLVGVPMKEYROELTILPYFOFATNPLPDNVAAR
PUBMED	21396507		ILPEROGAWDTGLVMTSEFRMDNRNRLIEGSLGRDLATAAGTHRAFAARSVRKLPFYIG
REFERENCE	11481430		DFRFEHWDGRIGMTTNNLPAHVLAPNVYSISYNGRGTAPGTVFGRALARHVTGDT
AUTHORS	2 (bases 1 to 323450)		SAIPLAETVPVDPWPRTLSAFYHAGAQAHHFIDKRF"
TITLE	Gouzy J.		complement(3024..3740)
JOURNAL	Direct Submission		/gene="SMc01579"
	Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO		/complement(3024..3740)
COMMENT	EU Consortium		/gene="SMc01579"
	MELILO EU Consortium:		/function="miscellaneous; hypothetical/global homology"
	Laboratoire de Biologie Moléculaire des Relations		/note="Product confidence : hypothetical
	Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,		Gene name confidence : hypothetical
	France, Laboratoire de Genetique et Developpement UMR6061-CNRS,		predicted by Codon_usage
	Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,		predicted by Homology
	France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,		predicted by Framed"
	Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,		/codon_start=1
	D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,		/evidence-not_experimental
	Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,		/transl_table=11
	B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte		/product="CONSERVED HYPOTHETICAL PROTEIN"
	des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,		/protein_id="CAC46789.1"
	B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr		/db_xref="GI:15075233"
	http://sequence.toulouse.inra.fr/meliloti.html.		/db_xref="SPTREMBL:O92NJ2"
FEATURES	Location/Qualifiers		/translation="MAVITFANAKGGAGKTTAALILSTELARQGNRVVVLDADPQRWI
source	1..323450		TSWSEVSRVANLEVISHTVPASLPCHIRELKGADFIIDLAKDAIIVALALGLSD
	/organism="Sinorhizobium meliloti"		HVLTPVQGCAMDARGAVQIILELIRHIGEKARINHSVLTNRVSLVTTALQTIKAL
	/strain="1021"		LASRGVSLDTPIVERVAYREIFECCGTLQMDPNRVSLDKARENAYALAAAEVQNLL
	/db_xref="taxon:382"		PVTARRALMSRLRSALPRAA"
gene	155..1519		complement(3949..5094)
	/gene="SMc01575"		/gene="SMc01582"
CDS	155..1519		complement(3949..5094)
	/gene="SMc01575"		/gene="SMc01582"
	/function="miscellaneous; hypothetical/partial homology"		/EC_number="1.1.1.-"
	/note="Product confidence : hypothetical		/function="small molecule metabolism"
	Gene name confidence : hypothetical		/note="Product confidence : putative
	predicted by Codon_usage		Gene name confidence : hypothetical
	predicted by Homology		predicted by Codon_usage
	predicted by Framed"		predicted by Homology
	/codon_start=1		predicted by Framed"
	/evidence-not_experimental		/codon_start=1
	/transl_table=11		/evidence-not_experimental
	/product="HYPOTHETICAL PROTEIN"		/transl_table=11
	/protein_id="CAC46787.1"		/product="PUTATIVE ALCOHOL DEHYDROGENASE PROTEIN"
	/db_xref="GI:15075231"		/protein_id="CAC46790.1"
	/db_xref="SPTREMBL:O92K34"		/db_xref="GI:15075234"
	/translation="MTPATRTGLSLFAALALAMELAPANAQDAYGGYGGGVMLVT		/db_xref="SPTREMBL:O92NJ1"
	PEGDILDYIPGEAEVHAMDRGRRTVLVDPMGNIVATVPNDGVRGQRREYGRDGY		/translation="MTITANKSVPTAVKFCAGRIKELADHCKALGTKKPLLVTDRGLA
	PEPRDRGYGSEPGCTGAIPYRDIAAPAPEREDLPNSLPSLSDREAAAYDPQYDDP		PMATIQALDILEAGGLGRAIFADVPDNPDRNLNLAGVKAFRGGHDCGVVAFGGGSL
	LAQPMPTVTCKSRAEIAALQVFLDRGFSFGVIDGKGSNVTXAEIAWQATGET		DLCKVAFMAGQTRPVWDFEDIGDMMWTRASVEGLAPIVAVPTTAGTGEVGRASVITN
	LDNNPEDIILRFRNGGLITTYITIAADAGFFVASIPEDTAHRAQLPHLSFTSVT		SASHVKVITFPKPLPGVTICDPELTVGMPKVITAGTMDAFACILEAYSPPYHPMS
	EMLGEKPHMDEALRELNPGVDFSIPGTTIKVNPNGPKKGVARIADVADKARKOVLAY		AGIALGEMRLVKEYLPRAYKDGLDLARANMMSAAAMGAVAFQKGLAIHSLHPVGA
	DEAGKIIAAYPTSTIGSSDTPSPSGTVHERIAFDPPGYTPNPKINFOQGANDRILOLQP		IYNTHGMTNAVVPVPLRFRNSAIEEKIGRAAYILGAGGFGDYVLRLEELGV
	GPNGPVTVMIALSKPTYIGHGTPEPSKIGKTSQHGCVRLTNWDATDELGMKMWSTGVTV		PKLSALGVGCTDRIDEMAENAIVDPTAGGNPVELTLDAAEKLFACET"
	EFVD"		complement(5091..6476)
gene	1707..2996		/gene="SMc01588"
	/gene="SMc01576"		complement(5091..6476)
CDS	1707..2996		/gene="SMc01588"
	/gene="SMc01576"		/EC_number="1.2.1.-"
	/function="small molecule metabolism"		/function="small molecule metabolism"
	/note="Product confidence : putative		/note="Product confidence : putative
	Gene name confidence : hypothetical		Gene name confidence : hypothetical
	predicted by Codon_usage		predicted by Codon_usage
	predicted by Homology		predicted by Homology
	predicted by Framed"		predicted by Framed"
	/codon_start=1		/codon_start=1
	/evidence-not_experimental		/evidence-not_experimental
	/transl_table=11		/transl_table=11
	/product="PUTATIVE OXIDOREDUCTASE PROTEIN"		/product="PUTATIVE ALDEHYDE DEHYDROGENASE PROTEIN"
	/protein_id="CAC46788.1"		/protein_id="CAC46791.1"
	/db_xref="GI:15075232"		/db_xref="GI:15075235"
	/db_xref="SPTREMBL:O92NJ3"		/db_xref="SPTREMBL:O92NJ0"
	/translation="MATERALPNLWHATAPAPRTAPLAGDLTVEVAIVGGFTGLSA		/translation="MTMIRCVSPVDGEVYAEPAIPLEMARQAVAHARLAQKAWARRP
	ALHLAETGIRTAIVIEARMIGFGSGRNVGLNVMVQPPDLDLIATIGAAAGNRLLOEL		LDERVKLVLAGVARNLMEVDEVPQWQMRPVRYGGEKFGFNEKRSNYSAAADAL
	GDGPFVYDLVARKHGECEAVRNGTLHMSVAGBLKEIREAGQWKNRGAPVEQVLSAE		KPLVVEESDRFERRIAREPHGVVFIAPNYPYMTAINTVAPALMAGNTVILKHSOT

Chen, Y., Woo, L., Kitajima, J. P., Okura, V. K., Almeida Jr., N. F., Zhou, Y., Boyce Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J. A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.

TITLE

Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

FEATURES

source

Location/Qualifiers

1. 11164
/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"

/strain="C58"

/db_xref="taxon:180835"

94. 1578

/gene="mtlK"

/note="Atu4451"

94. 1578

/gene="mtlK"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="mannitol 2-dehydrogenase"

/protein_id="AAL45245.1"

/db_xref="GI:17742929"

/translation="MTCKLSLATLDEAKTAAPAYSRNDSAGIVHFGVGNFHRAHQ
AVYLLDFNTGAHDFAIIGAGVLPDSATWREKLAQDFTTVVEODNNRTGARVGP
MIDLKVDGTQAITDITADPKIRIVSMTITEGGVFIDAGSENPQHPAIAEDCKNPSA
PKTVFGEIVAGLAKRDKDLQPTVMSCDNIPIHGKVTYKNAVVGSLAALSPAPANWIG
ENAFPNMVDRIPTATGERERNIARDQFDGDNWPFVEEFKQWMEHDFNAPGRPAL
EKAGQVQKVPYAEHMKIRILNGHAAIYAPALDIFHVHEAMEHPLKIFLAFLAKLE
KLEIPIPVPTDLNDYGLFERFLNPKIGDTIPRLAQDGSNRQPKFILPSTLDR
LSRGEDIVLSLSALMCRYFYGTSDSGKEIVFNDSAERLQAAAIKAKDDPVAFAL
DDIFGEVAESALFRKRESHALTTLWQEGTAKTLQLYLDLDDKLAEK"

1585. 2304

/gene="Atu4452"

/note="Atu4452"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL45246.1"

/db_xref="GI:17742930"

/translation="MRDGTGPLLIFDCGVLDSEPVSIIVLDMLSHLGVTWCEEA
YERFLGRSVASMTTLEEGVETDIDFLOHMRATLEPERTELKPIDGLAEITDELL
PLKRVASSQPERIRYSLGTLGIDKFEPVFSATVRNGKPAFDILFLAARSMGVE
PROQVIEDSPAGIAAAGAGMGVFAFTGGSHARFPAPREKTAGLADGAVDAMPDLV
OLVGSYVGRGFDGQVERDAG"

2359. 3936

/gene="Atu4453"

/note="Atu4453"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="ribitol kinase"

/protein_id="AAL45247.1"

/db_xref="GI:17742931"

/translation="MMRQNEVAVDVGTASARAGIFDPAGRLARSIHPILMQRPRENH
AEHSDTIWNAVCTIAALADAGVLPQSIATAGFDTACSLVTRDERGEPVSVSTDD
DRDPTIIVLDRHAIGEADRLTAGSHRVLDFAAGNSVSPQMOMPKMLWLTHTMPOSWRM
SEAFDLADFLTWATGSAQRSNCTQTAKNFWLAQHPGWQADYLAFLAGLDDLERAGL
PETVMPGAGIPLSPAAAEALGLDTCGQVAAGMIDAYAGALGALGGLAEADVGKHYA
LIATGSSCLVAMSTQPMGRSLMGYPYQWAVLPGLHWLVEGGQSATGALLDHIVRMHAAG

gene

CDS

GEPTALHARIVARVTELRLEGEAFADRLHLVLPDFHGNRSPADPHAVGVVSGLTLD
TSRDSLCRLYRTAIALGARHVLDMERFQAVETLHVTGGHVKNPILMEIYADV
KTDVDPATADAVLLGTAMTAATAGVHVASLAAAGAAAMPYGNAEISGNPALAAHYERD
YRFLAMRYRHOELES"

Complement(4041. .4874)

/gene="Atu4454"

complement(4041. .4874)

/gene="Atu4454"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="transcriptional regulator, Rpir family"

/protein_id="AAL45248.1"

/db_xref="GI:17742932"

/translation="MVEKLLNLTRAMKACTPSPERKTAKYLIEHDELPEPTAOTLAA
KLSLSPMTVGRFLRSIGYRQFSDIRADLRHAEETVAGDQVTAEGGERHSNPFQSLLS
QQIOAIOATAFDMTTQTITWRLAMNETTSADDVCLATSPGDLGADRHFFGRULEEYGH
YLGSDNAAVALMDYDPARTLLVIMCGNLLPLQRLSTTARKGGYTLITTRFYEW
GPESVDICLAMPOSQNGGQILQIVLSLEFTLCALSAGADDAGKARVKNLAALKRSLR
S"

4897. 5067

/gene="Atu4455"

4897. 5067

/gene="Atu4455"

/note="putative; ORF located using Glimmer"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAL45249.1"

/db_xref="GI:17742933"

/translation="MNGDAFARPCRKANHDLTIVETSPFGHVLHRTGMRIIVDGLLYFR
AIFYSRRIIPAV"

5170. 5907

/gene="Atu4456"

5170. 5907

/gene="Atu4456"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL45250.1"

/db_xref="GI:17742934"

/translation="MIPDAARLAFGNLFAAFTRSVFWKVLGTLTLLVLAALWFAIRSF
IYFALPWIDTLIPGVDPAGNLSFVFAIFAGIGLALALALLISPTAVIAGLFLDDVA
DVEKQNPDDPGKAMPIGEAVLSSIRKFGVITAGNLVALLLLLVPGVNLIAFLFLVN
GLLGRFFFEFAAMRFRSPADARLFRVKHRTTVMAGLVIAAFIAVLPFLNLLTFLFAA
SLMVHLHKAVSRDRDPSFAGGTQLRG"

complement(5869. .6579)

/gene="Atu4457"

complement(5869. .6579)

/gene="Atu4457"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="endonuclease III"

/protein_id="AAL45251.1"

/db_xref="GI:17742935"

/translation="MRVKATYSKDELTEIFRRFSIORPEPKGELEHTNPFTLLVAVAL
SAQATDVGVRNTRALFKVADTPKMLALGEEQLIGHIKTIGLYRNKAKNVIALSOML
IDNFGVGVNTRRELVTLPGVGRKTANVMMSMAFGVPTLAVDTHVFRINRCLAPCK
TTDEVEDRLVRIIPEOYLFLHAHHWLIHLGRYCCKARKPECERCVIADICKSPEKTFDI
PAPLVELPAQLFGAAGE"

6689. 7171

/gene="Atu4458"

6689. 7171

/gene="Atu4458"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11


```

/product="AGR_L_810p"
/protein_id="AAK88977.1"
/db_xref="GI:15158761"
/translation="MTDDSSERYRAPALDKGLDILELIARTDGLTGLEIAKATGKSP
NELVRLDLRVGRYVORLDGDRFSLTKMFLAHFAPITRLYSFAAPVMRDPFSASA
EOACHLAVDRGCVVTAQOBSATYWMKSMVGVGAOMSLFHTGSGHILLAFQTEAOREI
MTPEQVSGSTVPADLAEURQVRNSGFESMDSLOTAGVRNISAPILTLGGTALAVI
KCPITTLGGTAPREFCEQLIRDAAKRISEVV7GNV"
complement(972..1937)
/gene="AGR_L_812"
complement(972..1937)
/gene="AGR_L_812"
/translation="MTSKLEQLRAITTVADTGDIKAVARLKPVDCTNPTIVLKALG
TPAFADAVKEAVANGRGQSDAVAAVADRLAISVGAALAGLVPGRVSTVEDVDLS
PDQASITKARHITASYKRGIERILIKLASTWEGIKAAEVLOSIGIDCNLTLLFS
QAOAIAACAAKAFILISPPVGRILDMYKSTGETYTAETDPCGVSVRSIYAYKANGIG
TVVNGASFRNVGEALAGCDRLTISPALLEELDKNGKLERKLSPNVRAEALQSLD
EKAFRWAMNEMATEKLSEGLIRFAKDLVTLREMVRKELTAAAA"
2133..3164
/gene="AGR_L_814"
2133..3164
/gene="AGR_L_814"
/translation="AP001509) transcriptional regulator"
/codon_start=1
/transl_table=11
/product="AGR_L_814p"
/protein_id="AAK88979.1"
/db_xref="GI:15158763"
/translation="MAKTAVAKLRRETHITYSEASLSRLRAWLYNNGMGTQKDVABRL
GISSTVIRLLDEAMKSEVQIWINIEGIDFVSLAGOLEAYGDEAVITIPSPGPSA
GPSTACVSLGTTAAGTAKAGVAGLGOFLSEVVPNGATIGVWGRTMTASLSFRPP
RRENCVYSLGGTVAVHOTNPLDITWRLASALGAEYMFELAPLLVSVEVKRALIEK
CGLATLDLAETLAVISCDGIDGPHSTISLSEGIISRETRELVDAGCVCDTMEFNFD
AEGRSVDHPINORAMAIDLDTLRKAKHIVLASGGAHRAIAIRATIKRIGCWTLTDEA
AARALMELV"
complement(3161..3448)
/gene="AGR_L_815"
complement(3161..3448)
/gene="AGR_L_815"
/translation="HYPOTHETICAL 11.0 KD PROTEIN_Y4KP"
/codon_start=1
/transl_table=11
/product="AGR_L_815p"
/protein_id="AAK88980.1"
/db_xref="GI:15158764"
/translation="MRLIWTFRYLRELDIAIGVYTAERNPRAAAKVVRAIHOTTARLLS
DNPHLGRSGEGTRELVPGLPIVAYRVTEEQVDIVFVQHAAREWPDPA"
complement(3454..3690)
/gene="AGR_L_816"
complement(3454..3690)
/gene="AGR_L_816"
/translation="HYPOTHETICAL protein"
/codon_start=1
/transl_table=11
/product="AGR_L_816p"
/protein_id="AAK88981.1"
/db_xref="GI:15158765"
/translation="MKSTIELSDDIKRRDLTAERNSTPSRIIEDALSHGSRSLAWQE
KWTSGVRAGLAADAGEFVTEEDINLVKYAKA"
complement(3743..4615)
/gene="AGR_L_818"
complement(3743..4615)
/gene="AGR_L_818"
/translation="O6-methylguanine-DNA methyltransferase PA2118
[imported] - Pseudomonas aeruginosa (strain PA01)"
/codon_start=1

```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

```

/transl_table=11
/product="AGR_L_818p"
/protein_id="AAK88982.1"
/db_xref="GI:15158766"
/translation="MNANIMLNEDIPTIGSDYDVRGVIELLTLDYREOPSLAEIAAR
LGQSPLOLQKTFTRWAGLSKPAFLQAVTLDHAKLLREEDLPLETSTSEVGMSPGRLL
HDLFVTEAMSPGEMKARGGLTIRYGFHASPFGALAVMTIDRGLAGCAFADPGDERA
CFEDMACRPNADVEDREATAFYAARIFEPAWMTADKPLRVLLGLTDFQYRVKSLLL
KIPMGRAVTVYSNIACDICOPTASRAVGAAGANPVSFVVPCHRAVGSKALTGYHWGL
TKRAMLGWETGKA"
complement(4882..5364)
/gene="AGR_L_819"
complement(4882..5364)
/gene="AGR_L_819"
/translation="HYPOTHETICAL protein"
/codon_start=1
/transl_table=11
/product="AGR_L_819p"
/protein_id="AAK88983.1"
/db_xref="GI:15158767"
/translation="MDTLNDQPIFAAELVPHRSRGRGFRLLILITGTFACLTYGGFFL
ATGAMPICLFGDLFLLYIAFRANYRAAKAREVSVSRNLISIRKFSAGRMVYERF
NPFWARVRVRHDEIGIVSMHVTGEGRATDICSFLNDDRESFAKATGGALATVRRRM
"
5402..6184
/gene="AGR_L_820"
5402..6184
/gene="AGR_L_820"
/translation="endonuclease III PA3495 [imported] - Pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_820p"
/protein_id="AAK88984.1"
/db_xref="GI:15158768"
/translation="MTVAKKRSSTQTLKKSIPAORRPPVRVKTAYSKDELTEIFRRFS
IQRPKEGIEHTNPTFTLLVAVALSAQATDGVNRRATRAKFKVADTPEKMLALGEEQL
IGHTTGLVRYNKAKNVIALSOMLIDNFGVEPKTREELVPLGVGRKTKANVMSMAF
VGLPTADTVHFRIANRCLAPKMTDEVEDRLVRIIPEOYLPHAHHLWILLHGYCKK
ARKPECRCVIADICKSPEKTFDIPALVELPAQLFGAAGE"
complement(6146..6883)
/gene="AGR_L_822"
complement(6146..6883)
/gene="AGR_L_822"
/translation="HYPOTHETICAL protein"
/codon_start=1
/transl_table=11
/product="AGR_L_822p"
/protein_id="AAK88985.1"
/db_xref="GI:15158769"
/translation="MIFDAARLAFGLNFAAETRSVFWKVLGLTLLVLAALWFATRSIF
IYFALPWIDTLIPGVPDWAGWLSFVFAIFAGIGLALALLISPTAVIAGLFLDDVA
DVPEKQNPDDPGKAMPICEAVLSSIKFEGVVIAGNLVALLLLVPCVNLIAFELVN
GYLLGRFFFEFAAMRFRSPADARLFRVKNHRTTVFMAGLVIAAFVLAFLNLLTLPFAA
SLMVHLHKAVSRDRPSPFAAGTEQLRG"
7179..8012
/gene="AGR_L_824"
7179..8012
/gene="AGR_L_824"
/translation="HYPOTHETICAL protein"
/codon_start=1
/transl_table=11
/product="AGR_L_824p"
/protein_id="AAK88986.1"
/db_xref="GI:15158770"
/translation="MVEKLLNLNUTRAMKAGTPSERKIAKYLIEHLDELPEETAOTLAA
KLISLSPMTVGRFLSLGYROFSDIRADLRHAEETVGADQPVTAEGGERHSPPFQSLLS
QIOAIQTAFDMTQTITWRLANNEITADDVCATSPGDLADAGRHFGRLLECRGHIH
YLGSDAAVYVALWDYDITLIMDCGNNLLPIQRLSTWARKGYRTLIITTFYEW
GPESVDLCAMPQSQNGQGLLQVLSLLEFTLCALSAGADDAGKARVKVKNLAALKRSLR
S"
complement(8117..9694)
/gene="AGR_L_826"

```

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS	complement(8117..9694) /gene="AGR_L_826" /note="D-ribulokinase (EC 2.7.1.47) - Klebsiella pneumoniae" /transl_table=11 /codon_start=1 /product="AGR_L_826p" /protein_id="AAK88987.1" /db_xref="GI:15158771" /translation="MMRNLVAVDVGTASARAGIFDPAGRLIARSIHPILMQRPRENH AEHSDIWNACVIAADALADAVLPQSTAAIGFDATCSLIVIRDERGEPVSVSTDD DFDTIIVLHDRAITGEADRLTASGRHVSPEMQMQLMWLTKTHMPSWSRM SFADLADFTWKTGAQSNCTQTAKWNEFAQHPQWQADYLAFAGLDLDLKRAGL PETVMPGGATGLPSPAEALGLDTCQVAAGMDAYAGALGALGCAEDVGVKHA LIAGTSSCLVAMTGPGRSLWQVLPQHWLVEGGSATGALLDHTVVRHAA GEPDIALHVARIVTLELEGEAFADRLHVLDPFHGNSPLADPHAVGVVSLTLD TSPDSLRLYRTAIALGARHVLDMERGYAVETLHVTGGHVKHKNPLMELYADVT GKRIVFATDAVLGTAMTAAATAGGVHAAAGAAAMPYGNAEISNPALAAHYERD
Query Match	6.3%; Score 259.2; DB 1; Length 14945;
Best Local Similarity	54.3%; Pred. No. 3.7e-42;
Matches	591; Conservative 0; Mismatches 488; Indels 9; Gaps 3;
QY	581 GGCTCCCCCTATGACATCGACGGGATCAACGCTGGGATCTGTCATTTGGGTAGGTAA 640
DB	11906 GGTTCGGGCTATTCCCGGAACGATCTGTCCGCGGCGATCTGTCATTTGGCGCTTGGCAA 11847
QY	641 CTTTTTTCGACCGATGAGCGTCTACGTCGACGAGATCTTGAACACGCTCGGACTG 700
DB	11846 TTTCCACCGCCCATCAGCGGGTCTATCTTGATGACCTGTTCACACCGGTGCGGACCA 11787
QY	701 GCGCATTTGTTGGTCTGACGGGCGAGTGACCGTTTCAAGAAAAAGCCGAGGAATT 760
DB	11786 CGACTTCGCCATCATCGCGCGCGGCTTCTGCCCTTCGATGCCAGTCCGCGAAACT 11727
QY	761 CAAGCCACGACTGCTGTATTCCCTGACCGACGAGCGTCCGTCGCGACAGACGACGGT 820
DB	11726 TCCGCGACAGGATTTCCTGACCACTGTCTGTCGACGAGCACAATAATCGCACCGGTGC --- 11670
QY	821 CGCGTCATGGCGCGTCTGCTACTATCTCTTCCCGCGCGATCGGGAAGCGGTCT 880
DB	11669 CGCGTCACCGGGCGGATGATCGACATCTCAAGT---TGGCATACCCAGGCCATCAT 11613
QY	881 GAAGCATCTTGTGATCGCGGCATCCGATCGTTTCCATGACGATCAGGAAGCGGCTA 940
DB	11612 CGATACGCTTCCGATCCGAAATCCGATCGTTTCGATGACGATTACCGAGGCGGTTA 11553
QY	941 CAACATCAACGACGACCGTGGCTTCGATCTGGAGATCGGCGAGTAAAGGCGGACCT 1000
DB	11552 TTTCAT---CGATGCTCTCGGCTCTTCAATCCGACGATCCGGCCATTCGCGAAGACGG 11496
QY	1001 CAAGAACCGGAAAGCGCTACCGTCTTTCGGTTACGTGTGCGAGGCGCTCGCTCGTCG 1060
DB	11495 CAAGNACCGTCCGCTCGAARACCGTTCGGTTTCATCTGTCGGGCTGAAGGCGG 11436
QY	1061 TTGGGATCGCGGTGGTAAGGATTTACGGTTCATGCTGTCGATACCTGCGTCAATACGG 1120
DB	11435 GCGGATAAGGACCTGACGCGCTTACCGTCTATGCTGCGACACATTCGCCACAATGG 11376
QY	1121 CANTGTCGCCGACGACCTTCTCGGCTATCGAAGCGCGCGATCCGGAGTTGGCGAA 1180
DB	11375 CAAGGTGACAAAACCGGGTGGTGGGCTGCGAGCGCTCTCCGACCGCGCTTGGCAA 11316
QY	1181 GTGATTCAGGAAACGCGACCTTCCCGAAGGATGTTTCATCGCATCACCCCGACCGT 1240
DB	11315 CTGNATCGGGGAAATATGTCGCTTTCGGAATTCATGTGTCACCGCATCACCCCGGCAC 11256
QY	1241 TTCGCGGAAATCGCCAAAGAGCTCAACGCGCGCGAGTGGGTGGATGACGACCTGCGCGCT 1300
DB	11255 CGGCGAGGCTGAACGCAATATCGCCGCTGACCATTTTCGGCATCGAGGATAATTGCGCGGT 11196
QY	1301 GGTGCCGAGGATTTCCATCAGTGGTCTCGAAGACCATGTTTGGGATGCGCTCGCC 1360
DB	97474256

Db	11195 CTTCTGCCAAGAGTTCAAGCAATGGTGTATGGAGGATAATTTCCCGCAGCGCTCGGGC 11136
QY	1361 GCTTGAARAAACCGCGCTGCAGATGGTCGGGAGCTGACGAGCTGGGAGTACGTCAAGAT 1420
Db	11135 GCTCGAARAAACCGCGCTGCAGATGGTCAAGAGTGTTCGGCCCTTACGAGCACATGAAGT 11076
QY	1421 CCGAATGCTCAATCGAGGGGATGTCTGCTCTCCAGGCATTTCTGTCGGCTATGA 1480
Db	11075 CCGCATCTGAATGGTGGCCAGCGCAATCGCTATCCCGCAGCGCTGCTCGACATCA 11016
QY	1481 GAATGCTGATGACCCCATTTGAAGACAGGAACTCTTGGCAATCTGAAGAATCTATCTCAA 1540
Db	11015 TTTGCTGATGACGCGGATGGAGCATCCGCTGATCCGGCGGTTTCTGGCCAAAGCTCGAAA 10956
QY	1541 CAAGGATGTATCCCGACCTTGAAGGCGCTTTCAGGATGACGCTCGAAGGCTATCGGA 1600
Db	10955 AGACGATCATCCCGTTCATTCGCGGTACCTTGACACCGACCTCAACGACTATTTCCG 10896
QY	1601 CAGCTCATCAGCGCTTTCTCCAAAGGGCATCTCGGACAGACGCTCCGGATTTGCTAG 1660
Db	10895 GCTCATCGAGCGCGCTTTCTCTCAACCCGAAGATTGGCGATACCATTCGCGCGCTCGCGCA 10836
QY	1661 CGATGGCT 1668
Db	10835 GGATGGTT 10828

RESULT 12	AF018073	9810 bp	DNA	linear	BCT 22-OCT-1997
AF018073					
LOCUS					
DEFINITION	Rhodobacter sphaeroides operon regulator (smoc), periplasmic sorbitol-binding protein (smoE), sorbitol/mannitol transport inner membrane protein (smoF), sorbitol/mannitol transport inner membrane protein (smoG), sorbitol/mannitol transport ATP-binding transport protein (smoK), sorbitol dehydrogenase (smoS), mannitol dehydrogenase (mtkI), and periplasmic mannitol-binding protein (smoM) genes, complete cds.				
ACCESSION	AF018073.1 GI:2338757				
VERSION	AF018073.1				
KEYWORDS	Rhodobacter sphaeroides.				
SOURCE	Rhodobacter sphaeroides.				
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.				
REFERENCE	1 (bases 6163 to 8431)				
AUTHORS	Schneider,K.H., Giffhorn,F. and Kaplan,S.				
TITLE	Cloning, nucleotide sequence and characterization of the mannitol dehydrogenase gene from Rhodobacter sphaeroides				
JOURNAL	J. Gen. Microbiol. 139 (Pt 10), 2475-2484 (1993)				
MEDLINE	94075966				
PUBMED	8254318				
REFERENCE	2 (bases 5865 to 6635)				
AUTHORS	Schneider,K.H. and Giffhorn,F.				
TITLE	Polyol metabolism of Rhodobacter sphaeroides: biochemical characterization of a short-chain sorbitol dehydrogenase				
JOURNAL	Microbiology 141 (Pt 8), 1857-1863 (1995)				
MEDLINE	96032011				
PUBMED	7551049				
REFERENCE	3 (bases 6679 to 8212)				
AUTHORS	Schaefer,A., Stein,M.A., Schneider,K.H. and Giffhorn,F.				
TITLE	Mannitol dehydrogenase from Rhodobacter sphaeroides S14: subcloning, overexpression in Escherichia coli and characterization of the recombinant enzyme				
JOURNAL	Appl. Microbiol. Biotechnol. 48 (1), 47-52 (1997)				
MEDLINE	97419532				
PUBMED	9274047				
REFERENCE	4 (bases 4855 to 6750)				
AUTHORS	Stein,M.A., Schaefer,A. and Giffhorn,F.				
TITLE	Cloning, nucleotide sequence, and overexpression of smoS, a component of a novel operon encoding an ABC transporter and polyol dehydrogenases of Rhodobacter sphaeroides S14				
JOURNAL	J. Bacteriol. 179 (20), 6335-6340 (1997)				
MEDLINE	97474256				

9335280
 PUBMED
 5 (bases 6163 to 8431)
 REFERENCE
 Schneider,K.H. and Giffhorn,F.
 TITLE
 Direct Submission
 Submitted (09-MAR-1994) Lehrstuhl fuer Angewandte Mikrobiologie,
 UNIVERSITAET des Saarlandes, Geb. 2, Saarbruecken D-66041, Germany
 6 (bases 4855 to 6162)
 REFERENCE
 Stein,M. and Giffhorn,F.
 TITLE
 Direct Submission
 Submitted (22-MAR-1996) Lehrstuhl fuer Angewandte Mikrobiologie,
 UNIVERSITAET des Saarlandes, Geb. 2, Saarbruecken D-66041, Germany
 7 (bases 1 to 9810)
 REFERENCE
 Stein,M. and Giffhorn,F.
 TITLE
 Direct Submission
 Submitted (11-AUG-1997) Lehrstuhl fuer Angewandte Mikrobiologie,
 UNIVERSITAET des Saarlandes, Geb. 2, Saarbruecken D-66041, Germany
 On Aug 21, 1997 this sequence version replaced gi.1236173.
 COMMENT
 FEATURES
 Location/Qualifiers
 1..9810
 /organism="Rhodobacter sphaeroides"
 /strain="Si-4"
 /db_xref="taxon:1063"
 681..1634
 /gene="smoc"
 681..1634
 /gene="smoc"
 /note="ABC transporter; putative; Smoc"
 /codon_start=1
 /transl_table=11
 /product="operon regulator"
 /protein_id="AAC45765.1"
 /db_xref="GI:2338758"
 /translation="MIDRPESEPTPLDARAGWLYVAGLTQDQIARELQTSRQAO
 RLVSRAISERLIHVRHSVGLHLEALLRRFGLKLARVAPSLGSEVDPLPSIAPT
 RAAEVERLRSERPVMVAEGTGRSLRATVEEMTSMVCQHKIVSLNGNISADGSASY
 DVIFRIARVAPHVMPMPVIAQDAARELFHALKPVQSVLRARNADVTYVGLGQM
 GEDAPLLKDGITPELHEMODLGAAGVAGVDFDESGREYLETSSINQRVAGRVRELSE
 DRTVVAIAGGRKKLAALHAGLRGLFNLITDEFTAQALLS"
 1663..3052
 /gene="smoe"
 1663..1668
 /gene="smoe"
 1687..1692
 /gene="smoe"
 1731..1735
 /gene="smoe"
 1742..3052
 /gene="smoe"
 /note="ABC transporter; putative; Smoe"
 /codon_start=1
 /transl_table=11
 /product="periplasmic sorbitol-binding protein"
 /protein_id="AAC45766.1"
 /db_xref="GI:2338759"
 /translation="MTAFRALMGACAAVALSSRAGATTIVAVNVNGDMTRMOGLMS
 EFNQAPDITVEWTVLEENLRQKVTITDIATKGGQFDVLTITGYEVPWKGQWLVS
 NDLPYDADDLLFAIRNGLTVDGELYAAPFGSSIMIMYRKDLMEKAGLTFMPDAPT
 DFVRAAQKMDKDAEYVYICLRGKAGGNNMFLSAMANSYGARWFENWQPDGE
 AWKATLDYLDWMTNYGPPGAKNGFNENLALFOQKCGKMWIDATVAASFVTPPEST
 VADKVGFLAPDPTGKFKANLGAWNLAIPAGSQKVDAAKQFIARATSKDYAEVLASK
 EGWAVPGRITSLYENPEYKVPFPAKMTLDSINRAADPTHPAIVDPVPTVGVQFVAIPE
 FQGIATVGQQFSAALGMSAEALQAQQQFTTREMTRAGYIK".
 3061..3088
 /note="smoe-smof intergenic region repeat"
 3102..3120
 /note="smoe-smof intergenic region repeat"
 3144..4025
 /gene="smof"
 3144..3147
 /gene="smof"
 /note="putative"
 3153..4025
 /gene="smof"
 /note="ABC transporter; putative; Smof"
 /codon_start=1
 /transl_table=11
 /product="sorbitol/mannitol transport inner membrane
 protein"
 /protein_id="AAC45767.1"
 /db_xref="GI:2338760"
 /translation="MATQHSKTAARLMISPAVILLFLMWIVPLSMYLSMTLYFSLRYNLLM
 PGMSEFTGMDNVYFTDPAFSAALNTILLVVGVLITVVGVLALLLLDQPPWQGG
 IVRVLVIAPEFVMPVTSALVKNMFMNVNGMFEAHIARGLGLPPDFLSQAPLSTIG
 IVANQWLFPATLILLTALQSLDREQMEAEAMDGASALDRFTHITVPHLTRAITVVVLI
 QTIFLLGVFAEILVTYNGGPGTASTNITLYLVAOSLLNYDVGGSGAGGIVAVVLANIY
 AIFLMRMIGKNLDA"
 4016..4857
 /gene="smog"
 4016..4020
 /gene="smog"
 /note="putative"
 4027..4857
 /gene="smog"
 /note="ABC transporter; putative; Smog"
 /codon_start=1
 /transl_table=11
 /product="sorbitol/mannitol transport inner membrane
 protein"
 /protein_id="AAC45768.1"
 /db_xref="GI:2338761"
 /translation="MSRRTSTRRLTIVTLAAWTIAFLIFFPILMTVLMSPKSEGDALK
 APPAMLFSDWTLSQYADVQERSNVARHEMNSVVISLSTLVALAIAIPAANAMAEVPG
 RTKDLVLMMLSTLKMMPAVGLIPLYLIFRDTGLDTRIGLIVLITLNLPIVVMVLY
 TYFKEIPEGEILEAARMGDATLGSEILYILTPMVPGIASLTLLNLYLWNAEAFWTLOL
 TTSRAAPLTQPIASYSSPEGLFYAKLSAASTMAIAPILILGWFSOKOLVRGLTFCGAVK
 ".
 4857..5863
 /gene="smok"
 4857..4860
 /gene="smok"
 /note="putative"
 4865..5863
 /gene="smok"
 /note="ABC transporter; putative; formerly polk; Smok"
 /codon_start=1
 /transl_table=11
 /product="sorbitol/mannitol transport ATP-binding
 transport protein"
 /protein_id="AAC45769.1"
 /db_xref="GI:2338762"
 /translation="MGKITLRNVOKRFGAEVVIPLSLDIEDGEVFEVFGSPGCGKST
 LLRLTAGLEDVSDGQIMIDGRDAYTEPPPAKRGKLAWFQSYALYPHMTYVKNIAFPLRM
 AKMEPQETERRVSNAAKILNLTNYLDRRPGQSGQQRQVRAIGRAIVREPAALFDEP
 LSNLDAALRVNMRLEITELHOSLETMTIYTHDOVEAMTADKIVLNAEIRIOVQSP
 LTLYRNANLFAVAGFIGSPKNLLEGPEAAKHGATTIGIRPEHIDLREACAGWEGVG
 VSEHLGSDTFLHVVHAGNPTLTVRTGGEFGVHGHDVWLTPOADKIHRRFGADGAL".
 5852..6635
 /gene="smos"
 5852..5859
 /gene="smos"
 /note="putative"
 5865..6635
 /gene="smos"
 /EC_number="1.1.1.14"
 /note="short-chain alcohol dehydrogenase; formerly pols;
 Smos"
 /codon_start=1
 /transl_table=11
 /product="sorbitol dehydrogenase"
 /protein_id="AAC45770.1"
 /db_xref="GI:2338763"
 /translation="WRLDGKTLITGSARGIGRAFAEAYVREGARVALADINLEAARA
 TAAETGPACAIALDVTDOASTDCRVAELLDWMSIDILVNNALFDLAPLVEITRES
 YDRLEPAINVSGTLFMMQAVARAMIAGRGKKLIINWASOAGRGALGVGYCATKAAPF
 SLTOSAGLNLIRHGINATPAIPVVDGHDWDGDAKFADYENLPRGEKKRQVGAAPF
 GRMGRAEDLTGMAIFLATPEADYIVAQTYNVNDGGNNNS"

```
gene 6680..8124
/ gene="mtik"
RBS 6680..6684
/ gene="mtik"
/ note="putative"
CDS 6691..8124
/ gene="mtik"
/ EC_number="1.1.1.67"
/ note="Mlik"
/ codon_start=1

Query Match 6.0%; Score 247.4; DB 1; Length 9810;
Best Local Similarity 53.4%; Pred. No. 9.4e-40;
Matches 590; Conservative 0; Mismatches 506; Indels 9; Gaps 3;

QY 584 TCCCCCTATACATCAGCGGATCAAGCCCTGGGATCGTGATTCGGTGTAGTAACATT 643
|||||
DB 6711 TCCCTCTATAGACCGCAAGCGCTCACTCCGGGATCGTCCATATCGCGGTCGCAACTT 6770

QY 644 TTTTCGAGCCCATCAGCGGTTCTACGTCGAGCAGATCTTGAACACGCTCCGGACTGGC 703
|||||
DB 6771 CCACGGGGGATCAGCGGCTCTATCTCGAGATCTTCTCGCGTGGCGAGGCCACGA 6830

QY 704 GATTGTGGTGTGGCTTACGGGACGTGACCGTTCAAAGAAAAAGCCGAGGAATCAA 763
|||||
DB 6831 CTGGGCCATCTCGCGCGCGGCTCCGCCACCGATGCGCGGATGCGCGAGGCTCTGGC 6890

QY 764 GCCCAGGACTGCTGATTCCTTGACCGAGACGGCTCCGTCGGCAGACACGGTGGC 823
|||||
DB 6891 CGCGAGGACAAATCTCTGACGGGTGATCGAGCTCGATCCGGCGGGCCA---CCGGGCGC 6947

QY 824 CGTCATGGCGCGCTGCTGATCTATCTGCTCCCGCGCGATCGGAAGCCGCTGCTGAA 883
|||||
DB 6948 CAGGTGGGGGATGGTGGCTTCTCGCGTCCGAGCCACAAATCGGCGCTGATCGA 7007

QY 884 GCATCTTGTGATCCGGCCATCCGATCGTGGTTCATGACATCAGGAAGCGCGCTACAA 943
|||||
DB 7008 GGCCATGTCGGATCCGCGCATCCGATCGTCTGCTGACCGTACCGAGGCGGCTATTA 7057

QY 944 CATCAACGAGACGCGGTGCTGATCTCGAGATCGGACGATCGGACGTAAGACCGACCTCA 1003
|||||
DB 7068 TGTGATGCTCG---GGCGCTTCGATCCGACGATCCCGATATCGTGGCGGATCGGC 7124

QY 1004 GAACCCGAAAGCGCTTACCGTTTCGTTACGTGGTGGAGCCCTCGCTCGTGGTTG 1063
|||||
DB 7125 CCATCTCGCGCGCGGACCGCTTCGCGCGATCTCTCGCGCCCTCCCGCGCGCGCG 7184

QY 1064 GATGCGCGTGGTAAGGATTTACGGTATGTCCTGTCATACCTGCTCATACGCGCA 1123
|||||
DB 7185 CGACGCGGGGTTACACCTTCACCGTATCTCTCGACAACTCCCGCGCAACGGCCA 7244

QY 1124 TGTGCGCGCAAGGCTTCTCGGCTATCGAAGCGCGGATCGGAGTTGGCGAAGTG 1183
|||||
DB 7245 TGTACCCGCAACGCCGTGGTGGGCTGGCGAGCTCTACGACGCGAGCTTGGGGCTG 7304

QY 1184 GATTGAGAAACCGACCTTCCCGACGGAATGTTGATCGATCAACCCGACGCGTTTC 1243
|||||
DB 7305 GGTGAAGCGGAGGTGGCTTCCCGAAGCGCATGTCACCGCATCACCCCGCGCACCGG 7364

QY 1244 GCGGAATCGCAGAGATCAACCGCGGCGAGTGGGCTGATGACGACCTGCCGCTGGT 1303
|||||
DB 7365 CCGCA---CGAGCGCAACTGGCGAGGGTTCGGGCTCGCGATCGGTCGGGTGCAC 7421

QY 1304 GCGCGAGGATTTCCATAGTGGTCTCGAAGACCACTTTCGGATGCGGCTCGCGCGCT 1363
|||||
DB 7422 CTGCGAGCGGTTCCGGAGTGGGTGATCGAGATCATTTCCCGCGGACGCGCGCT 7481

QY 1364 TGAAGAACCGCGCTGACAGTGGTGGGACGTGACGAGCTGGGAGTACGTCAGATCCG 1423
|||||
DB 7482 CGAGAAGTGGCGGTGACCTTCCACCGCGCATGTCATGCTACGAGGCGATGAAGATCCG 7541

QY 1424 ATGCTCAATGAGGCGATGTCATGCTGCTGCCAGGCACTTCGTCGCTGATGAGAA 1483
|||||
```

```
Db 7542 CATCCTGAACGGGGGCATCGGTTGATCGCTATCCCTCGGCGCTCATGGACATCCAGCT 7601
QY 1484 TGTGGATGACGCCATGAAGACAGCGAACTCTTGGCAATCTGAAGAACTATCTCAACAA 1543
|||||
Db 7602 CGTGCAGCGGCATGCCCATCGCTGATCGGCGCTTCTCTGCACAGGTCGAGGTCGA 7661
QY 1544 GGATGTCATCCGACCCCTGAAGGCGCTTTCAGGATGACGCTCGAAGGCTATCCGGACAG 1603
|||||
Db 7662 GGAGATCCTGCCCATGTCGCGCGCTGCGCGACACAGCATCCCGACTATCTTACCCCT 7721
QY 1604 CGTCATCAGCGTTCCTCCACAGGCGATGTCGACACAGAGCTCCGCGATGTCAGCA 1663
|||||
Db 7722 GATCGAGAGCGCTTCTCGAACCCGAGATCGCGACACAGCGAGGCTCTCCCTCGA 7781
QY 1664 TGGCTGTTCCAAAGGTTTCAGGTGTTTC 1688
Db 7782 CGGTTCGAACCGGACCGGAAGTTC 7806

RESULT 13
LOCUS AE005663
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 282
of 290.
ACCESSION AE005663
VERSION AE005663.1
KEYWORDS GI:12519330
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933.
REFERENCE 1 (bases 1 to 10955)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL Nature 409 (6819), 529-533 (2001)
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 10955)
AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Source
1. 10955
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
complement(118..1461)
/gene="gntP"
/note="Z5919"
complement(118..1461)
/gene="gntP"
/function="transport; Transport of small molecules:
Carbohydrates; organic acids, alcohols"
/ note="Residues 1 to 447 of 447 are 100.00 pct identical
to residues 1 to 447 of 447 from Escherichia coli K-12
Strain MG1655: B4321"
/transl_table=11
/codon_start=1
/product="glucuronate transport system permease 3"
/protein_id="AAG59503.1"
```


[illegible]


```

/db_xref="GI:15141287"
/db_xref="SPTREMBL:Q92TU6"
/translation="MMDGASNAAPRIIRGTPALPDLGSRVALITGSSGGGLTMAHALC
EAGSVILGRDEARLAGEALEEHCFTVGTSAFDVTKSAEIGRAVADILAEGRGID
ILVNNAGIORTPLHEPPEDAFRVLETNLTSALFLVQAQVVOGMIEBRECTIINICSV
QSELARPSTAPYASKGGKMLTKGMALDWRGIRVNLGAPGYFKTELSALSVDK
FSTWLSORTPLGRWGDTEGELAAAVFLAASFSFVGHILYVDGGITSCL"
gene
CDS
complement(1063..1361)
/genes="Smb20693"
/genes="Smb20693"
/genes="Smb20693"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49800.1"
/db_xref="GI:15141288"
/db_xref="SPTREMBL:Q92TU5"
/translation="MPDYAASKGVGVOLAKAFATNGOPGTFRSMPLPPGYIATDTEA
SRNDPSRAESILARIIPRAAGAGPRTLPAARSEFWPRHPITSRASSPWTAAAGWA"
1508..1584
/genes="Smb20694"
/genes="Smb20694"
/genes="Smb20694"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49801.1"
/db_xref="GI:15141289"
/db_xref="SPTREMBL:Q92TU4"
/translation="MGADGDPMAALDARMVRVIERLRVADMSAVPDINAGNTNAPAT
MLGNRCAGFILGTA"
complement(1765..2271)
/genes="Smb20695"
complement(1765..2271)
/genes="Smb20695"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49802.1"
/db_xref="GI:15141290"
/db_xref="SPTREMBL:Q92TU3"
/translation="MTLLPVALGGTELVAWRLDQAVHAPTWDSGEGAYBVGGWNSKG
VRAYVCSLPATVAILLEVAVHKGFRALDMVAHNTAAADAGDVYVYDPSKVPNPWA
RCPGAGQAQVAGDDLRHRHREVAIPSAVSPHSWNLVFLGGAAAPALAKFOEPFALD
TRLHPGCT"
complement(2268..2753)
/genes="Smb20696"
complement(2268..2753)
/genes="Smb20696"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49803.1"
/db_xref="GI:15141291"
/db_xref="SPTREMBL:Q92TU2"
/translation="MALAKATPPAAGGGELOK IETLLGGSRYLAHLTNALDAHELL
LRGLPASADHLVGNLVIFIGKTESLEKAVGMSLRTWQRKTEPNKPLSQEQSGRAWKF

```

```

A
AETLVKATDIFGSQAEAEQWLERPAVGLDQRRP IDLLTTPPGVELVEDHLDRLLEYGIY
complement(2855..4549)
/genes="Smb20697"
complement(2855..4549)
/genes="Smb20697"
/function="MACROMOLECULE METABOLISM; Macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative peptidase, similar to arginine
utilization protein RocB"
/protein_id="CAC49804.1"
/db_xref="GI:15141292"
/db_xref="SPTREMBL:Q92TU1"
/translation="MTNAQEETVSRAGRSIDSERVQFALRMTSPSETGTGGEASFA
DRHLGGLGELPYFREHPDGLLASHGEPLTRNNVALVRGTGKRTLVMAGHFDVTSTD
NYHELKALACDSLALKDALTESLSARTGRSEERALQDLASGDFLPGRGLLDMKSL
AVAIACLEQFAADTDROGNMLVATPDEERESGRMSLRDALPCLVRDEIDETAGIN
LDVTSQDGGSGRAVYAGTICKLPLFALVIGCSSHASTPFEQVSAWAAGLLALE
GNASLADRDNDISPPPICLEAKDLRDGYEVTPFERFIATFNWLYASQMTADALFERF
EEVLTGANEAIETFAQSAEYKLVGRSAGLPKAPRLLSFOELRAAAARVFGHFD
FYAEKREFEAQSDNPLVATROLTENLVGRSLGPAIVIGFSGLHYPPSHRLAECND
RSLHOATEKARAGLNDPARSLIWKPHFYGISDMSFLGAASGVSDNTPI SRLVD
RSENALRFTVNLGPGWREFHQKFERVIEPYAFVLPVELVSEIARTFLCDDHRD"
complement(4701..6353)
/genes="Smb20698"
complement(4701..6353)
/genes="Smb20698"
/function="CELL PROCESSES; Transport of small molecules"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transport protein, similar to E. coli
multidrug resistance protein EmrB"
/protein_id="CAC49805.1"
/db_xref="GI:15141293"
/db_xref="SPTREMBL:Q92TU0"
/translation="MTIATATPSRIDSMAGLLVAGIVLATLEATASTVLSLRGDI
IGDYATPDEFALWDVGYTTFKLLIGMAASLWMSRDFPRNLVVGSTLAMGACGIAAM
TARLDLLVALRMIOGSPGGTLLVGGQAILFLTFPHSRQPLLOQAFAMGSVVAATTAP
ALQGLLLDSQSTWIFFSIIPALAAAGLLLLADGPVATRAQRPPEDWIGFSLISAAL
FCFTYVFSQGSRWDFEERILMWLVIGTATLAFLGQOVLLANGQGLDFTLTFEDF
CFATVSVAGAAALFGSAFLIPSAVSVLAFTPDAGQLLLPSGALFALLIAEFLM
QLRRVPPVATVPFGILIMAAWMLSGSTSESGADGMAAILIRGAGLGLFLISITLI
AFSNLSRNLASGILGNTGROLGILGYSAIQTLIEHNVSHNLAVLGANVTAGAPAV
ADRLTTAALLTAKGMDAAAASRGASLLRGAVAGSTVIAFDATFNAIGLLFVIAAP
VLVGKIGILARYAKARAEKGRAGVAAKPSCHHTLPIKRVAS"
complement(6350..7447)
/genes="Smb20699"
complement(6350..7447)
/genes="Smb20699"
/function="CELL PROCESSES; Transport of large molecules;
protein, peptide secretion"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative protein secretion protein, HlyD family,
similar to E. coli EmrK"
/protein_id="CAC49806.1"
/db_xref="GI:15141294"
/db_xref="SPTREMBL:Q92TU9"
/translation="MESLEKVAITGARGSGNCVAVARNRKKMIGLALCLGAVVVVV
AGWTWARESAASTNAYVRGDDVTSIAPKAVGYTAVEDEQAVRAGDVLFRIDRD
YRAKLAQAVANVAEAEARLTNVDAEMALQHALIRQAEQRSSVVAELNLAARAYRRR

```

ELINSETISQAHVDESDAKSRAEANVLAASATVEAQOORIVLAQAQREAAVAQA
EAARDLAGIDLESTVVRAPVGVGNRVRGRVAPGASLLDIVPLDNWIVANEKE

Query Match	5.78;	Score 236.2;	DB 1;	Length 189333;
Best Local Similarity	50.78;	Pred. No. 1.3e-37;		
Matches 655;	Conservative 0;	Mismatches 618;	Indels 18;	Gaps 3;
QY	539	GATTACCGCGGAAACCCCTTAAGTCTCTCTCTGCAATGTCCAGGCTCCCCCTCATGACAT	598	
DB	44081	CTTTTCGAGAAGACGATGACCCCTTCGCGCACCGTCAAGAGCCCATTTACGACCT	44140	
QY	599	CGACGGGATCAAGCCTCGGATTCGTCATTTTCGTTAGTAACTTTTTCGAGCCCATGA	658	
DB	44141	CGGACAGTCAACCGTCGCGATCGCCACCTCGGCATCGGTGCTTTCATCGCGGCACCA	44200	
QY	659	GGCTTTCTACGTCAGCAGATCTTGAACAGCTCCCGACTGGCGATTGTTGGTGTGG	718	
DB	44201	GGCGCTCTACACCATGGGTGCTCTCCGAGGATTCGCTCGGGGATCTCGCGCGTCTC	44260	
QY	719	CCTGACGGGAGTGACCGTTCAAGAAAGAAAGCCGAGGATTTCAAGCCCGACGACTGCCT	778	
DB	44261	GCTGGCAGCCCGAG-----ACCGGAGCGCGCTTCATCCGACGAGCGGCT	44308	
QY	779	GTATTCCCTGACGAGACGGTCCGTCGCGCAAGACACGGTGGCGTCAATGGCGCGCT	838	
DB	44309	CTACACCTCGCGCTCGCAGGCGGAGGCTC---CGAGCTCAGCGTGGTTCGGCAGTGT	44365	
QY	839	CGGTGACTATCTGCTTCCCGCGCGGATCCGGAAGCGGTGCTGAAGCATCTTGTGATCC	898	
DB	44366	CGTCGAGCTCTTTCGCGCGGAGACCGCGGAGCGGCTCTCGCGCATGGCGGATCC	44425	
QY	899	GGCCATCGCATCTGTTTCCATGACGATCAGGAAGGGGCTACAAACATCAACGAGACGAC	958	
DB	44426	CGGCACAGGATCGTGTGACGATCACCAGAGGGCTATTGCCATTAACCCGGCAAC	44485	
QY	959	CGGTGCGTTCGATCTGAGAAATGCGGAGTAAGGCGGACCTCAAGAACCCCGGAAAGCC	1018	
DB	44486	CGGCAGCTCGACGAAGGCCACCGGACATCGTCCAGCACTCGCAAAACCCCGCGCGCC	44545	
QY	1019	GTCTACCGTTTTCGTTACGTGTGTCGAGGCGCTCGGTGCTGTTGGGATCCGGTGTAA	1078	
DB	44546	CGCATCTGCGATCGGGTTCATCTGCGAGGGGATTTCCGCGCGCGCTCAGCGCGGTATCGC	44605	
QY	1079	GGCATTTACGGTCATGCTGTGATAACCTGCTGATCAACGGCAATGTCGCCGCCAAGGC	1138	
DB	44606	ACCTTCACGCTGCTCTCTCGATATCTTCCGGGNAATGGCCATGTCCTGAGCGGTAT	44665	
QY	1139	CTTCCCTCGGCTATCGGAAGCGCGGATCGGAGTTGGGAGTGGGATTTGAGGAAACGC	1198	
DB	44666	CGTCACGCGATTCTCGGAAGCGCGATCCCGCCTCGCGCGCTCGTCCGCAACGTCGC	44725	
QY	1199	GACCTTCCCGAAGATGTTGATCGCATCACCCGACCGTTTCGGCGGAAATCGCCAA	1258	
DB	44726	GTCCGCTTCGAC---ATGGTGGATTCGATCTTCCCGGCCACACCCGACGCGCGGAG	44782	
QY	1259	GAAGCTCAACCGCGCGAGTGGGCTGATGACGACCTGCGGTGGTGGCGGAGGATTTCCA	1318	
DB	44783	CGCGTCCGCTCGCAATGGGCTCGAAGATGCTGGCCGATCATGACCGAACCTTCGG	44842	
QY	1319	TCAGTGGGTGGAAGACAGTTTGGGATGGCGCTCCCGCGCTTGAAAAGCCGCGT	1378	
DB	44843	GCAATGGGTGATGAGGAGGATTTCCCGCTCGCGCGCGCGCTTGGGAAAGCGCGGAGC	44902	
QY	1379	GCAGATGGTGGGACGTCGAGGATAGCTCAAGATCGAATGCTCAATGCAGG	1438	
DB	44903	GCTCTTCGTCACAGCTCTCCGCTTCGAGTTTCATGAGCTCCGGCTTCGACGCGGAG	44962	
QY	1439	GCATGTATGCTGCTTCCAGGCAATTCGTGGTGGGTATGAGAAATGTGATGACGCCAT	1498	
DB	44963	CCATTCCACGCTGCTATCTCGGCTATCTCGCGGTGCGGAGACCGTGGCGGACGCCAT	45022	
QY	1499	TGAAGACAGCGAATCTTGGCAATCTGAAGAACTATCTCAACAGGATGTCATCCCGAC	1558	

Search completed: May 18, 2003, 23:45:56
Job time : 11490 secs

Db	45023	GGCACTTGGCGGATGGAAGCGTTGCTCGAAGGTTGATCGGCCACGAGGTGAGCCCCAC	45082	
QY	1559	CTTGAAGGCGCCTTTCAGGATGACGCTCGAAGGCTATCGGACAGCGTTCATCAGCGGTTT	1618	
DB	45083	CCTTCGGAACCTTCGCGGCTTCGACCTTCGCGCTATCGGCGGAACTCCTTCAGCGCTT	45142	
QY	1619	CTCCAAAGGCGATGTCGACACGACGCTCCGATTTGCTAGCGATGGCTGTTTCCAAGGT	1678	
DB	45143	TCGCAATCCTGCGCTTTCGCGCACCGACCTGGCAGATCGCCATGACGCGCTCGCAGAAGCT	45202	
QY	1679	TCAGGTGTTCTGACGGAACACCGTGGTGGGGGATCGAAGACAAGGGGACCTGTACG	1738	
DB	45203	GGCGACGCGCTCCTCGGCTCGATCCGCGACCGCTTTCGAGGAGGGGCGGCTATACCG	45262	
QY	1739	TATAGCTTCGGAATTCATCTCTCGAAATGCTGCGTGGTTCGCGACGAGAAAGCGCG	1798	
DB	45263	GCTGGCGCTCGGGTTTCGCGCTTCGCGCTATGCGCGGCTGCGCGGCTGACGAGCGGCGC	45322	
QY	1799	GACGTATGATCGTCCGAGCGGCTTATGGC	1829	
DB	45323	TCCCATCGAGCTCGCGCATCCGCATGCTGOC	45353	

DR F-F3DB; AAB23172.

XX Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
PT culturing its gene-transformed host cells, useful for producing
PT L-sorboside or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
XX acid production
PS Claim 9; Page 60-63; 72pp; Japanese.
XX
CC The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
CC (SLDH; AAB23172) and to the gene encoding it (AAA97430). SLDH has a
CC molecular weight of about 54 kDa and catalyses the conversion of
CC D-sorbitol into L-sorboside using NADP+ as a co-enzyme. SLDH specifically
CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
CC act on xylitol, ribitol, inositol and glycerol. The invention also
CC encompasses expression vectors and host cells comprising the
CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
CC The invention further relates to a method for preparing L-sorboside by
CC contacting the recombinant SLDH with D-sorbitol; a process for producing
CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
CC dehydrogenase and/or sorbosone dehydrogenase with L-sorboside; and a
CC process for preparing L-ascorbic acid or its alkaline earth metals salts
CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
CC L-sorboside or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
CC acid production. The present sequence represents the Gluconobacter
CC oxydans SLDH gene.
XX
SQ Sequence 4115 BP; 833 A; 1133 C; 1270 G; 879 T; 0 other;
Query Match 100.0%; Score 4115; DB 21; Length 4115;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGATGCTCCAGGTGACACTAGAGGATCCGGTTTGGCAGCGCTCCCTAGATT 60
DB 1 AAGCTTGATGCTCCAGGTGACACTAGAGGATCCGGTTTGGCAGCGCTCCCTAGATT 60
QY 61 GATCGGGCTGTGTTGACCGACATGATGCTGTGTGGACGTGCCATTTGCGAGCGGGCGTGC 120
DB 61 GATCGGGCTGTGTTGACCGACATGATGCTGTGTGGACGTGCCATTTGCGAGCGGGCGTGC 120
QY 121 GACCGGGACACAGCGCTGCTGCTTTTACAAAGGGGCTGAGTCAATGCGGTGGGCT 180
DB 121 GACCGGGACACAGCGCTGCTGCTTTTACAAAGGGGCTGAGTCAATGCGGTGGGCT 180
QY 181 GGCACATAGTTGCGAAGAGCAGTTCGCGGCAAGCAGACAGCATGAACAGCAGTCCGA 240
DB 181 GGCACATAGTTGCGAAGAGCAGTTCGCGGCAAGCAGACAGCATGAACAGCAGTCCGA 240
QY 241 AGACGAGGAATCTCTCGGCTCTACCGGATTTGAAGAGCAGACCCCTCTGAGATCGG 300
DB 241 AGACGAGGAATCTCTCGGCTCTACCGGATTTGAAGAGCAGACCCCTCTGAGATCGG 300
QY 301 TTTTGTGATGCTCCGTTCCGGAGGATCTCGAACGGGCTGTGGGGTCTCATGCGTTC 360
DB 301 TTTTGTGATGCTCCGTTCCGGAGGATCTCGAACGGGCTGTGGGGTCTCATGCGTTC 360
QY 361 TGATCGGATGCGCAAAAGTCTCTGAACAGGACGCTCCCGGGAGGCGAGTTCAGAGTTC 420
DB 361 TGATCGGATGCGCAAAAGTCTCTGAACAGGACGCTCCCGGGAGGCGAGTTCAGAGTTC 420
QY 421 GAAATGGCTCTGTTGAAACCGTCAATCGGTTTACGTTTTCGGGGCTATGATGCA 480
DB 421 GAAATGGCTCTGTTGAAACCGTCAATCGGTTTACGTTTTCGGGGCTATGATGCA 480
QY 481 CATGCGCGGCTTGTGGTCCCGTCCAGCAGCGCCGCAACACAGGAGATTCATGA 540
DB 481 CATGCGCGGCTTGTGGTCCCGTCCAGCAGCGCCGCAACACAGGAGATTCATGA 540
QY 541 TTACGCGGGAACCCCTTAAGTCTCTTCTGCAATGTCCAGGCTCCCGCTTATGACATCG 600
DB 541 TTACGCGGGAACCCCTTAAGTCTCTTCTGCAATGTCCAGGCTCCCGCTTATGACATCG 600
QY 601 ACGGATCAAGCCTGGGATCGTGCATTTTCGGTGTAGGTAACTTTTTCGAGCCCATGAGG 660

DB 601 ACGGATCAAGCCTGGGATCGTGCATTTTCGGTGTAGGTAACTTTTTCGAGCCCATGAGG 660
QY 661 CGTTCTACGTGCGAGCAGATTTCTTAACACACGCTCGGACTGGCGCATTTGTTGCTTTCGCC 720
DB 661 CGTTCTACGTGCGAGCAGATTTCTTAACACGCTCGGACTGGCGCATTTGTTGCTTTCGCC 720
QY 721 TGACGGGACGTGACCGCTTCAAAAGAAAAGCCGAGGAATTTCAAGGCCCGACGACTGCTGT 780
DB 721 TGACGGGACGTGACCGCTTCAAAAGAAAAGCCGAGGAATTTCAAGGCCCGACGACTGCTGT 780
QY 781 ATTCCCTGACGAGACGGCTCCGTCGCGCAAGACGCGTGCCTGATGCGGCGGCTGC 840
DB 781 ATTCCCTGACGAGACGGCTCCGTCGCGCAAGACGCGTGCCTGATGCGGCGGCTGC 840
QY 841 GTGACTATCTGCTTCCCGGCGGATCGGGAAGCCGCTGCTGAAGCATCTTCTTCATCCGG 900
DB 841 GTGACTATCTGCTTCCCGGCGGATCGGGAAGCCGCTGCTGAAGCATCTTCTTCATCCGG 900
QY 901 CCATCCGACATCGTTTCCATGACGATCAGGAAGCGGCTACAACATCAACGAGACGACCG 960
DB 901 CCATCCGACATCGTTTCCATGACGATCAGGAAGCGGCTACAACATCAACGAGACGACCG 960
QY 961 GTGCGTTCGATCTGGAAGATCGGCGAGTAAAGGCGGACCTCAAGAACCCGGAAGCGGT 1020
DB 961 GTGCGTTCGATCTGGAAGATCGGCGAGTAAAGGCGGACCTCAAGAACCCGGAAGCGGT 1020
QY 1021 CTACCGTTTTCGTTACGTGCTGAGCGCCCTGCTGCTGTTGGGATGCGGCTGTAAG 1080
DB 1021 CTACCGTTTTCGTTACGTGCTGAGCGCCCTGCTGCTGTTGGGATGCGGCTGTAAG 1080
QY 1081 CATTTACGGTCTGCTGATTAACCTGCTCATACAGCGCAATGTCGCCCGCAAGGCT 1140
DB 1081 CATTTACGGTCTGCTGATTAACCTGCTCATACAGCGCAATGTCGCCCGCAAGGCT 1140
QY 1141 TCCTCGGCTATGCGAAGCGCGGATCGGAGTTGGGAGTGGGAGTGGAGAAACGCGGA 1200
DB 1141 TCCTCGGCTATGCGAAGCGCGGATCGGAGTTGGGAGTGGGAGTGGAGAAACGCGGA 1200
QY 1201 CTTTCCGCAACGGAATGTTGATCGCATCAACCGCGGCTTTCGCGGAAATGCCAAGA 1260
DB 1201 CTTTCCGCAACGGAATGTTGATCGCATCAACCGCGGCTTTCGCGGAAATGCCAAGA 1260
QY 1261 AGCTCAACGCGGCGAGTGGGCTGGATGACGACCTGCGGCTGGTGGCGGAGGATTTCCATC 1320
DB 1261 AGCTCAACGCGGCGAGTGGGCTGGATGACGACCTGCGGCTGGTGGCGGAGGATTTCCATC 1320
QY 1321 AGTGGTCTGGAAGACCGAGTTTGGGATGCGGCTGCGGCTTGAAGAACCGCGGCTGC 1380
DB 1321 AGTGGTCTGGAAGACCGAGTTTGGGATGCGGCTGCGGCTTGAAGAACCGCGGCTGC 1380
QY 1381 AGATGGTGGGCGAGTGGGCTGAGTACGATCAAGATCCGATGCTCAATGCAAGGC 1440
DB 1381 AGATGGTGGGCGAGTGGGCTGAGTACGATCAAGATCCGATGCTCAATGCAAGGC 1440
QY 1441 ATGTCTGCTGCTGCTTCCAGGATTTCTGTCGGCTATGAGAAATGTGATGACGCGCATTC 1500
DB 1441 ATGTCTGCTGCTGCTTCCAGGATTTCTGTCGGCTATGAGAAATGTGATGACGCGCATTC 1500
QY 1501 AAGACACGGAATCTTGGCAATCTGAAGAACTATCTCAACAAAGGATGTCATCCGAGCC 1560
DB 1501 AAGACACGGAATCTTGGCAATCTGAAGAACTATCTCAACAAAGGATGTCATCCGAGCC 1560
QY 1561 TGAAGGGGCTTTCAGGATGACGCTCGAAGGCTATCGGAGACGGGCTCATAGCGGTTTCT 1620
DB 1561 TGAAGGGGCTTTCAGGATGACGCTCGAAGGCTATCGGAGACGGGCTCATAGCGGTTTCT 1620
QY 1621 CCAACAAGGCGATGTGCGACACAGCTTCGCGATTTGCTAGCGATGGGTGTTTCCAGGTTC 1680
DB 1621 CCAACAAGGCGATGTGCGACACAGCTTCGCGATTTGCTAGCGATGGGTGTTTCCAGGTTC 1680
QY 1681 AGGTGTTCTGACGGAACCGTGGGCGATTCGAGAACGCGGACCTGTCACGTA 1740

Db 1681 AGGTGTTCTGGACGGAACCGTGCTCGGGCGGATCGAAGACAAAGCGGACCTGTACGTA 1740
Qy 1741 TAGCGTTGGAATTCATCTCTCGAAATGCTCGTGGTCCGACAGAAAGCGCGGA 1800
Db 1741 TAGCGTTGGAATTCATCTCTCGAAATGCTCGTGGTCCGACAGAAAGCGCGGA 1800
Qy 1801 CGTATGAATTCGTCGAGCGGACTTATGGGAGCGCCGAATGGAAGTTGGCCAAAGCGGACG 1860
Db 1801 CCGATGAATTCGTCGAGCGGACTTATGGGAGCGCCGAATGGAAGTTGGCCAAAGCGGACG 1860
Qy 1861 ACTTCGAAGCTCTCTGAAGCTCCCGCGTTCGATGGTGGCGCATCTGGATAGCTCG 1920
Db 1861 ACTTCGAAGCTCTCTGAAGCTCCCGCGTTCGATGGTGGCGCATCTGGATAGCTCG 1920
Qy 1921 AACTGGATCAAAAGTCTATCGTCTGCGGAAGATCATCCGCAAAAGCGCTAAAAGCG 1980
Db 1921 AACTGGATCAAAAGTCTATCGTCTGCGGAAGATCATCCGCAAAAGCGCTAAAAGCG 1980
Qy 1981 CCATCCCGCCTGAATTCGGCTTTTAGGTAGCGACTGAACAGAAAAACCGGCTCTGGA 2040
Db 1981 CCATCCCGCCTGAATTCGGCTTTTAGGTAGCGACTGAACAGAAAAACCGGCTCTGGA 2040
Qy 2041 AGGACGCGGTTTTTTTATGCTCAGATCTGTCCCATCAGGACAGGATCAGGACCA 2100
Db 2041 AGGACGCGGTTTTTTTATGCTCAGATCTGTCCCATCAGGACAGGATCAGGACCA 2100
Qy 2101 CGATCAGCAGAACTCCGCTGGAGGGGAGCCCCATTCGAACTGTACGCCATGACGGCA 2160
Db 2101 CGATCAGCAGAACTCCGCTGGAGGGGAGCCCCATTCGAACTGTACGCCATGACGGCA 2160
Qy 2161 GCGACCCAGATCAGGATACAGAAAGATCAGTCCCATGGGACATCTCTTTCGCGGTT 2220
Db 2161 GCGACCCAGATCAGGATACAGAAAGATCAGTCCCATGGGACATCTCTTTCGCGGTT 2220
Qy 2221 GAGACTGTCTGTCTCGGCTGTCTAAAGTCTTAGGGGCGCAAGATCAAAAGC 2280
Db 2221 GAGACTGTCTGTCTCGGCTGTCTAAAGTCTTAGGGGCGCAAGATCAAAAGC 2280
Qy 2281 TGTGCGTCCGCTTAAATCCGCTCCCAAGCGGATGATGCGGGCCACCGCTCTGTCG 2340
Db 2281 TGTGCGTCCGCTTAAATCCGCTCCCAAGCGGATGATGCGGGCCACCGCTCTGTCG 2340
Qy 2341 CGTTTTCGCTCTGTCTGACATAGTTTCTGGGCCAGCAGCTCCGGATGATGTCGGG 2400
Db 2341 CGTTTTCGCTCTGTCTGACATAGTTTCTGGGCCAGCAGCTCCGGATGATGTCGGG 2400
Qy 2401 ATCAGGGTCCGACGACGCGGATTTCTGTCTCAGTTGCGCTGCGGGTATGCGCGAGA 2460
Db 2401 ATCAGGGTCCGACGACGCGGATTTCTGTCTCAGTTGCGCTGCGGGTATGCGCGAGA 2460
Qy 2461 ATACGATAGGCATCCGGCTCGTTCCGCTGGCGCGGATTTGTGCCGCTTTCGGCCCG 2520
Db 2461 ATACGATAGGCATCCGGCTCGTTCCGCTGGCGCGGATTTGTGCCGCTTTCGGCCCG 2520
Qy 2521 TCCCATGCTCTGGCGGACGCGCAATGCCCGGTGAACGCGCTGCAGAAAAATCGATTCC 2580
Db 2521 TCCCATGCTCTGGCGGACGCGCAATGCCCGGTGAACGCGCTGCAGAAAAATCGATTCC 2580
Qy 2581 TTCGGGTGAAGTCTCGCGCTGGGGCGGATCGGGCAGCGGATACGGAACAGTCCCGTC 2640
Db 2581 TTCGGGTGAAGTCTCGCGCTGGGGCGGATCGGGCAGCGGATACGGAACAGTCCCGTC 2640
Qy 2641 ATGAGGTTCTAAGCGGCGCGTATTATCGCATAGGCTTTGCCATTTTCGGGCGATAC 2700
Db 2641 ATGAGGTTCTAAGCGGCGCGTATTATCGCATAGGCTTTGCCATTTTCGGGCGATAC 2700
Qy 2701 ATCTCGAATTCGCTCGGTCGGCGGCGGATCGAAGACGATGCCGACTTCCTTGTG 2760
Db 2701 ATCTCGAATTCGCTCGGTCGGCGGCGGATCGAAGACGATGCCGACTTCCTTGTG 2760
Qy 2761 TTATCGGGGGAACTGGAAGCAGCTTTGAAGCGTTGATTTCTGTCGGTTCACCGG 2820
Db 2761 TTATCGGGGGAACTGGAAGCAGCTTTGAAGCGTTGATTTCTGTCGGTTCACCGG 2820

Qy 2821 CCGTCGATCTTTCGCGAGCTTCGCGCACAGGCAACAAGCGCGATGGCGTAAAGCTGATCT 2880
Db 2821 CCGTCGATCTTTCGCGAGCTTCGCGCACAGGCAACAAGCGCGATGGCGTAAAGCTGATCT 2880
Qy 2881 CGTTTTCGCGAGCGCGCAGCAATCTTGGCAGCGCGCAAAAAGCGCGCTGTGGGATCG 2940
Db 2881 CGTTTTCGCGAGCGCGCAGCAATCTTGGCAGCGCGCAAAAAGCGCGCTGTGGGATCG 2940
Qy 2941 GGAGCGCATTTTCGCGGGAAGCGCTCACTCAGCGCGCGCTTGAGGGCTTGAGTAGCGAA 3000
Db 2941 GGAGCGCATTTTCGCGGGAAGCGCTCACTCAGCGCGCGCTTGAGGGCTTGAGTAGCGAA 3000
Qy 3001 CCGTTATCGCGGATGCGCCAGCGCTGCGCCCATCAGTCTCCGAAAGGACCAACACC 3060
Db 3001 CCGTTATCGCGGATGCGCCAGCGCTGCGCCCATCAGTCTCCGAAAGGACCAACACC 3060
Qy 3061 GCGAAGCGCGGACACCAACATCTTGGCCAGATAGCATGTCAATCAACCTAGCAC 3120
Db 3061 GCGAAGCGCGGACACCAACATCTTGGCCAGATAGCATGTCAATCAACCTAGCAC 3120
Qy 3121 GCCGCTCACAGCGCAATGACAGATCAGAGCTAGGTAGGTGCTGATGCGCCAAACC 3180
Db 3121 GCCGCTCACAGCGCAATGACAGATCAGAGCTAGGTAGGTGCTGATGCGCCAAACC 3180
Qy 3181 GCCGCGCTTGGGTGTGTAGAGCTAGGAGTTACGAACTTATCGCTGTCTCATGCTTT 3240
Db 3181 GCCGCGCTTGGGTGTGTAGAGCTAGGAGTTACGAACTTATCGCTGTCTCATGCTTT 3240
Qy 3241 TGAGCGCAGGTTCTTCTGTTCTCATGAGGATATTTTATGCCACCTTGATCCAG 3300
Db 3241 TGAGCGCAGGTTCTTCTGTTCTCATGAGGATATTTTATGCCACCTTGATCCAG 3300
Qy 3301 ACTGCTACTTCGATCCCTTCCGCTCTGATCAGCAACTGATGATCTTTTATCAAGCT 3360
Db 3301 ACTGCTACTTCGATCCCTTCCGCTCTGATCAGCAACTGATGATCTTTTATCAAGCT 3360
Qy 3361 CTGCAATGTGGCTGCAGAAAGTGTGAACTGGTTGCGGGAAGCGGATCAATAATGGGTT 3420
Db 3361 CTGCAATGTGGCTGCAGAAAGTGTGAACTGGTTGCGGGAAGCGGATCAATAATGGGTT 3420
Qy 3421 CGGATTCGCGGGGCGTCTGTTATGCTGGCGGCGTCTGTCATCTGCCCTGTTCTG 3480
Db 3421 CGGATTCGCGGGGCGTCTGTTATGCTGGCGGCGTCTGTCATCTGCCCTGTTCTG 3480
Qy 3481 GGTCTGTGATGCTGCCGCTCGCGTGTGTTGCTGCGCAGGATATTCGCTTCTCCGT 3540
Db 3481 GGTCTGTGATGCTGCCGCTCGCGTGTGTTGCTGCGCAGGATATTCGCTTCTCCGT 3540
Qy 3541 CGCTTTCAGGCGCGCTCTTGGCTGGATCGAACTCAACATCCGATTTGGCTGGGCTT 3600
Db 3541 CGCTTTCAGGCGCGCTCTTGGCTGGATCGAACTCAACATCCGATTTGGCTGGGCTT 3600
Qy 3601 CCGCGAAAAAGCGGACAGGCTAACCGTTGCTGCGAGCTGTTCTGAAGATGTCAGT 3660
Db 3601 CCGCGAAAAAGCGGACAGGCTAACCGTTGCTGCGAGCTGTTCTGAAGATGTCAGT 3660
Qy 3661 GCTCAACCCCGCAGGGCTGAAGCAGTGGCGCTCTGCTGCTGCGCGGATCGAGAGA 3720
Db 3661 GCTCAACCCCGCAGGGCTGAAGCAGTGGCGCTCTGCTGCTGCGCGGATCGAGAGA 3720
Qy 3721 GCCACAGAGACGCAAGCTCTGCTGCGGACTGCGGCTATCGGCTCCAGTATAGCCAG 3780
Db 3721 GCCACAGAGACGCAAGCTCTGCTGCGGACTGCGGCTATCGGCTCCAGTATAGCCAG 3780
Qy 3781 AACTCGGGTTCAGTGCACGAGCTCCGCTGTCTGACAGAGACAGCTCGGTTGACG 3840
Db 3781 AACTCGGGTTCAGTGCACGAGCTCCGCTGTCTGACAGAGACAGCTCGGTTGACG 3840
Qy 3841 AGATCAGTCAATTCGGTTGTTCTCAAGCGCTTCAAGCCCATTTGTCGGGTTTCGGAAA 3900
Db 3841 AGATCAGTCAATTCGGTTGTTCTCAAGCGCTTCAAGCCCATTTGTCGGGTTTCGGAAA 3900

QY 3901 CATCAGGTCGGATCACTCAGCAGCTCCCGCGAGAGATATAAGCGACGGATCGGCCG 3960
 Db 3901 CATCAGGTCGGATCACTCAGCAGCTCCCGCGAGAGATATAAGCGACGGATCGGCCG 3960
 QY 3961 AGTTGCCGATCGCATCAGGACAGTTAGCTACGAAACCGGTTGCTCCAAATCCGTTTGACC 4020
 Db 3961 AGTTGCCGATCGCATCAGGACAGTTAGCTACGAAACCGGTTGCTCCAAATCCGTTTGACC 4020
 QY 4021 GGAGAGCCAGAAAAAGCTCCGGAATGCTGCATTATCCAGCCGACCCAGTTCTGCTGAGT 4080
 Db 4021 GGAGAGCCAGAAAAAGCTCCGGAATGCTGCATTATCCAGCCGACCCAGTTCTGCTGAGT 4080
 QY 4081 TTTGGTCAATCAGCTCCGGCGGGCGCTGAAAGCTT 4115
 Db 4081 TTTGGTCAATCAGCTCCGGCGGGCGCTGAAAGCTT 4115

RESULT 2

AA235672

ID AA235672 standard; DNA; 1458 BP.

XX

AC AA235672;

XX

DT 27-JAN-2000 (first entry)

XX

DE Gluconobacter suboxydans L-sorbose reductase nucleotide sequence;

XX

KW Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;

KW

mutant; ds.

XX

OS Gluconobacter suboxydans.

XX

FH Key Location/Qualifiers

FT CDS 1..1458

FT /*tag= a

FT /product= "L-sorbose reductase"

XX

PN AU9920390-A.

XX

PD 23-SEP-1999.

XX

PF 11-MAR-1999; 99AU-0020390.

XX

PR 13-MAR-1998; 98EP-0104546.

XX

PA (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX

PI Hoshino T, Tazoe M, Shinjoh M, Kon T;

XX

DR WPI; 1999-579276/49.

XX

P-PSDB; AAY49913.

XX

PT Set of genetically modified mutants not containing L sorbose reductase

XX

PS Example 4; Page 17-18; 33pp; English.

XX

CC The present invention describes a genetically engineered

CC microorganism derived from a microorganism belonging to the genus

CC Gluconobacter or Acetobacter which is characterised in that the

CC biological activity for reducing L-sorbose is substantially

CC nullified by gene recombination. The present sequence encodes

CC Gluconobacter suboxydans L-sorbose reductase.

CC N.B. This patent is equivalent to the basic NO9901197 in week 199949.

XX

SQ Sequence 1458 BP; 329 A; 386 C; 434 G; 309 T; 0 other;

XX

Query Match 23.2%; Score 955.6; DB 20; Length 1458;

Best Local Similarity 78.5%; Pred. No. 4.5e-241;

Matches 1144; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

XX

QY 537 ATGATTCAGCGCGAAACCTTAAGTCTCTCTGCAATGTCCAGGCTCCCGCCCTATGAC 596.

Db

1 ATGATTCAGCGCGAAACCTTAAGTCTCTCTCCCGCGGTGTGCAGGCTCCCGCCCTATGAC 60

QY 597 ATCCACGGGATCAAGCCTGGGATCGTGCATTTTCGGTGTAGGTAACTTTTTCGAGCCCAT 656
 Db 61 ATCAATGGGATCAAAACGGGGATCGTGCAITTTTGGCGTGGAAACTTCTTCCGGGCCCAT 120
 QY 657 GAGCGGTTCTACGTCGAGCAGAGATCTTGAACAGCCTCCGAGCTGGGGCAATGTTGGTGT 716
 Db 121 GAGCGTCTTCACTGTTGAACAGATCTTCAAGGAGATCCGAACTGGGGAATCATCGGCGTT 180
 QY 717 GGCTGACGGGCGAGTACCGTTCAAGAAAAAGCCGAGGAATTCAGAGCCAGGACTGC 776
 Db 181 GGTCTGAGGGGTAGCGHACAGGTCAAGAAGAAGCCGAGGAATTCAGAGACGAGGACTGC 240
 QY 777 CTGTATTCTCCCTGAGCCGAGAGCGGCTCCGTCGGCAAGAGCAGCGTCCGCGTCATGGGCGG 836
 Db 241 CTCTTTTCCCTGACCGAAACGGCTCCGTCGGCAAGAGCAGCGTTCGTGTTATGGGCGG 300
 QY 837 CTGCGTCACTATCTGCTTGGCCCGCGGATCCGGAAGCCGCTGCTGAAGCATCTTGTGAT 896
 Db 301 CTGAGGGATTACCTTTTGGCTCTCGCGATCCGGAAGCCGCTGCTGAAGCATCTCGCTGAC 360
 QY 897 CCGGCCATCCGATCGCTTCCATGACGATCAGGAAAGCGCGTACAAACATCAACGACGACG 956
 Db 361 CCGGGAATCCGTATCGTTTCCATGCAATCAGGAAAGCGGTTACACATTAAGCAGACG 420
 QY 957 ACCGTCGCTTCGATTCGGAGAAATGCGGAGTAAGGCGCCGACCTCAAGAACCCCGGAAAG 1016
 Db 421 ACAGGTGAGTTGATCTTGAGAACAGCGGTTTCAGCAGGATCTGAAGACACCCGAAACG 480
 QY 1017 CGCTCTACCGTTTTCGGTTAGTGTGAGGCCCTGCTGCTGTTGGGATGCCGTTGCT 1076
 Db 481 CCGTCCCAATCTTTGGATATGTTGGAAGGAGTGCGCCCGCGCTGACGAGGTGGC 540
 QY 1077 AAGGCATTTACGGTCACTGCTGTGATAACCTGCTCATACGCAATGTGCGCCGCAAG 1136
 Db 541 AAGSCCTTCAGGATCATGCTCTGCGATANTGCGGCATACGTAATGTGCGCCGCAAG 600
 QY 1137 GCCTTCCTCGGCTATGCGAAGCGCGGATCCGGAGTTGGGAAAGTGGATTGAGGAAC 1196
 Db 601 GCATTTCTGGGATACGCGAAGCGCGTGTCCGGAAGTGGCCAAAGTGGATTGAAGAAC 660
 QY 1197 GCGACCTTCCGCAACGGAATGTTGATCGCATCACCCGACCGTTTCGGGGAATCGCC 1256
 Db 661 GCGAGCTTCCCAATGCGATGTTGATCGCATACGCGGACCGTTCTGCTGACATGCG 720
 QY 1257 AAGAAGCTCAACGCGGCGAGTGGCTGATGACACCTGCGGCTGGTGGCGGAGATTC 1316
 Db 721 AAGAAGCTCAACGAAAGCCAGTGGCTGCACGACACCTGCGGCTGTCGAGAAGACTTT 780
 QY 1317 CATCAGTGGTCTGGAAGACCACTTTTGGGATGGCGTCCGCGCTTGAAGAACCGCGC 1376
 Db 781 CATCAGTGGTCTGGAAGACAGCTTTGCTGATGGCGGCTGCGCTGGAAGAACCGCGGA 840
 QY 1377 GTGCAGATGTTGCGGGGACGTGACGCACTGGGAGTACGTCAAGATCCGAATGCTCAATGCA 1436
 Db 841 GTGCAGTCTGTTGGGATGTGACGCACTACGAGCATGTAAATCCGATGCTGAATGCT 900
 QY 1437 GGGCATGTCATGCTCTCTTCCAGGCAATTCGTGCGGCTATGAGATGTGATGACGCC 1496
 Db 901 GGTCAATCATGCTCTGCTTCCCGGCTGTTCTGCGAGGATTTGAAATGTCGATCATGCC 960
 QY 1497 ATTGAACACAGCGAATCTTGGCAATCTGAGAACTATCTCAACAGGATGTCATCCCG 1556
 Db 961 CTTGCTGATCCCGATCTACGGCGTATCTTCGAGAACTTCTCGAACAAGAGCTCATCCCG 1020
 QY 1557 ACCCTGAAGGCGCTTCAGGCACTGACGCTCGAAGGCTATCGGGACAGCGTTCATCAGCGT 1616
 Db 1021 ACCCTGAAGGACCCCGGCGATCAGCTGGAAGGCTATCGGACAGCGTTCATCAGCGT 1080
 QY 1617 TTCTCCAAACAGGCGATGTCCGACGACGCGTCCGGATTTGCTACGATGGCTGTTCGAAG 1676
 Db 1081 TTCTCGAATCCGCGCATGTGGCGGATCAGACATTTGCGTATTTCCGGGACGCGGAGCTCGAAG 1140

Qy	1244	GGCGGAATCGCCCAAGAAGCTCAACGCGGCAGTGGGCTGGATGACGACTGCCGTGGT	1303
Db	7365	CCCCA---CGAGCGCAAACTGGCGCAGAGGCTTCGGCTCGCGGATCCGGTGCCCGTCA	7421
Qy	1304	GGCCGAGGATTTCATCAGTGGGTGTGGAAGACCAGTTTCCGATGGCGCGTCCGCGCT	1363
Db	7422	CTGGACCCGTTCCGGCAGTGGGTGATCGAGGATCATTTCCC CGCGGAGCCCGCGCT	7481
Qy	1364	TGAAAAAGCGCGGTGCAGATGTCGGGGACGTGACGGACTGGGAGTACGTCAAGATCCG	1423
Db	7482	CGAGAAGTGGCGTGACCTTCAACCCGCATGCTCATGCTACGAGSGCATGAAGATCCG	7541
Qy	1424	AATGCTCAATGCAGGGCATGTCATGCTCTGCTTCCAGGCATTTCTGGTCGGCTATGAGAA	1483
Db	7542	CATCTGAACGGGGGCCATCGGTGATGCGCTATCCGTCGGCGCTCATGGACATCCAGCT	7601
Qy	1484	TGTGGATGACGCCATTGAAGACACGCAACTCTCTTGGCAATCTGAAGAACATATCTCAACA	1543
Db	7602	CGTGACCGCGGCATGSCCCATCCGTGATCGCGGCCCTTCTTCACAAGSTCCAGTCCA	7661
Qy	1544	GGATGTCATCCGACCCCTGAAGGGCCTTCAGGATGACGCTCGAAGGCTATCGGACAG	1603
Db	7662	GGAGATCCTGCCCATATGTCGCCGCCGTGCGCCGACACCAAGCATCCCGACTATTTACCT	7721
Qy	1604	CGTCATCAGCCGTTTTCTCCAAACAAGGCGATGTCTGGACACACGCTCCGGAATTCAGCGA	1663
Db	7722	GATCGAGAGCGCTTCTCGAACCCCGAGATCGCGCACACGACGCGGAGGCTCTGCTCCGA	7781
Qy	1664	TGCGTGTTC AAGSTTCAGTGTTC	1688
Db	7782	CGGTTTCAACCGG CAGCCGAAGTTC	7806

RESULTS

RES001 J	
AAH65091	standard; DNA; 1509 BP.
ID	AAH65091 standard; DNA; 1509 BP.
XX	
AC	AAH65091;
XX	
DT	26-SEP-2001 (first entry)
XX	
DE	C glutamicum coding sequence fragment SEQ ID NO: 126.
XX	
KW	Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW	organic acid synthesis; ds.
XX	
OS	Corynebacterium glutamicum.
XX	
PN	EPI108790-A2.
XX	
PD	20-JUN-2001.
XX	
Pf	18-DEC-2000; 2000EP-0127688.
XX	
PR	16-DEC-1999; 99JP-0377484.
PR	07-APR-2000; 2000JP-0159162.
PR	03-AUG-2000; 2000JP-0280988.
XX	
PA	(KYOW) KYOWA HAKKO KOGYO KK.
XX	
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX	
DR	WPI; 2001-376931/40.
DR	P-PSDB; AAG89872.
XX	
PT	Novel polynucleotides derived from Coryneform bacteria, for identifying
PT	mutation point of a gene, measuring expression of a gene, analysing
PT	expression profile or pattern of a gene and identifying homologous gene
PT	
XX	
PS	Claim 8; SEQ ID NO: 126; 246pp + Sequence Listing; English.
XX	

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These are useful for identifying the mutation point of a gene derived from a mutant of coryneform bacterium, measuring expression amount and analysing the expression profile or expression pattern of a gene derived from Coryneform bacterium, and identifying a homologue of a gene derived from coryneform bacterium. Coryneform bacteria are useful for producing amino acids, nucleic acids, vitamins, saccharides and organic acids, particularly L-lysine. The present sequence is a nucleic acid described in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from the European Patent Office.

xx

Sequence 1509 BP: 341 A: 474 C: 376 G: 318 T: 0 other:

Db 393 GAACCTCTTGATCAAAAAATCATCGGATCCATCATTTGACTACGTGTTCCTCCCGAGGACC 452
QY 868 CGGAAGCCGTGCTGAAGCATCTTGTGTATCGGCGCATCCGATCGTGTTCCTCCATGACGATCA 927
Db 453 CAGCAGCGCGGTTCGAACCCCTCGCGAGGACTCCATCCGATTTTCCCTCAGGTTGA 512
QY 928 CGGAAGCGCGGTACAACTATCAACGAGACGACGGTGTTCGATCTGGAGATCGGGCAG 987
Db 513 CTGAAGCGCGGATACAACTATCGATCGCGCAGAGAGATTTCGACACACCAACCCCTCGAA 572
QY 988 TAAAGCGCGCATCAAGAACCGGAAAAAGCGTCT-----ACCGTTTTCG 1032
Db 573 TCGTTGCTGACCGGAACCCCTGACGGCGGGGATFACCTTCACATTTGCGAGACCTTCTTTG 632
QY 1033 GTTACGTGCTGAGGCCCTCGCTGCTGTTGGGATGCCGGTGGTAAGGCATTTTACGGTCA 1092
Db 633 GTTGTATCACTGCCGCAATTTATTTCCGAAAAAGATCAGGATCTACGCCATTTTACCATCA 692
QY 1093 TGTCTGTGATAACCTGCTCATACGCGCATGTGCGCCGGAAGCCCTTCCTCGGCTATG 1152
Db 693 TGAGCTGCGTAACATCCAAAGCAACGGCGATCTGGCTTAAGCGTTTCTTCCTCGCCTTCG 752
QY 1153 CGAAGCGCGCGATCCGAGTTGGCGAAGTGGATTGAGGAAAAACGCGACCTTCCCGAACG 1212
Db 753 CACATTCGTGCTTCTGAGCTCGCGGAATGGTGGAAAAACAGCTGGCCCTTCCCAACT 812
QY 1213 GAATGGTTGATCGCATACCCCGACCGCTTTCGGCGGAAATCGCCAAAGACTCAACGGCG 1272
Db 813 CCATGTGGACCGCATACCC---CTGAACACCGGACCGCGGATGACATCAAG 869
QY 1273 CCAGTGGCTGGATGACACACCTCGCGTGGTGGCGGAGGATTTTCATCAGTGGTGGTGG 1332
Db 870 AAATCGGCTACATCGATCGTGGCGAGTGGTTTCTGAAGATTTCAACCAATTTGGTCTCTCG 929
QY 1333 AAGACAGTTTTCGGATCGCGCTCGCGCTTGAAGAACCGGGTGCAGATGTCGGGG 1392
Db 930 AGGATGCTTTCACCCAGCGCGCCCGCTAGGAGGTTGGGTGCGTGTCTCCG 989
QY 1393 ACGTGACGACTGGGAGTACGATCAAGATCCGAATGCTCAATGCGAGGATGTCATGCTCT 1452
Db 990 ACGTGGAGCCTTATGAATTAATGAAGTGCCTGCTCAACGCTCCACACGAGGACTTT 1049
QY 1453 CTTTCCCGAGGATTTGTCGGCTATGAGATGAGATGAGATGAGCGCATTTCAACACAGCGAAC 1512
Db 1050 GCTACTTCGGCCACTTGGCTGGCCACACATGGTCCAGCAGCTATGCGCGATACCCGCT 1109
QY 1513 TCCTTGGCAATCTGAAGAACTATCTCAACAAAGGATGTCATCCGACCCCTGAAGCGCCTT 1572
Db 1110 TCCAGGATTTCTCTGCTTACATGGAGCGCGAGCCACCCCTACCCCTCAAGGAACCTC 1169
QY 1573 CAGGATGACGCTCGAAGGCTATCGGGACAGCGTCAATCAGCCGTTTCTTCCAAAGCGCA 1632
Db 1170 CAGGTGTGATCTAGATCTTATGACGCCAACTCATCGCGCGATTCGGCAACGCGCAG 1229
QY 1633 TGTCCGACACGCTCCGGATGCTACGATGGCTGTTCGAAGTTTCAGGTGTTCTGGA 1692
Db 1230 TCAAGACACCGTACCGCGCTGTGTGCGGAATCTTCGACCGGATTCCTCAAGTGGCTGT 1289
QY 1693 CGGAACACGCTCGTGGCGGATCAAGAACAGCGGACCTGTCAGTATAGCGTTTCGGAA 1752
Db 1290 TGCCAGTCTAGCGGAAAACTCGCAGCAGCGCGGAGCTCACATTTCTTCGACCCATCG 1349
QY 1753 TTGCATCTTCTCGAAATGTCGGTGTGTCGCGAGGAGGGC 1796
Db 1350 TCGCATCTCTGGCGCGCTACGCAAGGACCGACCGAGCAGGGC 1393

RESULT 7

AAH64966/c
ID AAH64966 standard; DNA; 349980 BP.
XX
AC
AAH64966;
XX

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 1.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizeuchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 7; SEQ ID NO: 1; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX Sequence 349980 BP; 79703 A; 91547 C; 98381 G; 80349 T; 0 other;

Query Match 5.2%; Score 213.6; DB 22; Length 349980;
Best Local Similarity 49.6%; Pred. No. 1.le-44;
Matches 617; Conservative 0; Mismatches 609; Indels 18; Gaps 2;

QY 568 CTGCGCAATGTCCAGGCTCCCGCTTATGACATCGACGGGATCAAGCCTCGGATCGCATT 627

Db 122407 CTTCGGAGTGCATATCCCGCTTCAACCGCGTACAGCTCGCCCGGCGATTGTCCACT 122348

QY 628 TCGGTGTAGTAACTTTTTCGAGCCCATGAGCGCTTCTAGCTGAGCAGATTTCTTGAAC 687

Db 122347 TCGGTGTGGCGGATTCATCGCGCTCACCAAGCGATGTACCTCAATCAATTGATGATG 122288

QY 688 ACGTCCGCGACTGGCGGATTTGTTGGTGTGCGCTGACGGGCGAGTACCGTTCAAGAAAA 747

Db 122287 AGGCAAGGCCCTTGGATTGGGCGATCATCGGCATGGTGTCTCCGATGTGCGCA 122228

QY 748 AAGCCGAGGAATTCAGGCCCGGAGGACTCCCTGATTCCCTACCGAGACGCTCCGTCG 807

Db 122227 TGGCGATGCGCTGGCCGAGCAAGATCACCTTTATACCTGACCACTAAAGCTCCTGTATG 122168

QY 808 GCAAGAGCACGGTGCCTCATGGCGCGCTGCGTGACTATCTGCTGCCCGCGCGATC 867

Db 122167 GAATCTTGATCAAAAAATCATCGGATCCATCATTTGACTGTTCTCCGTCGCGAGGACC 122108

Qy	868	CGGAAGCCGTGCTGAAGCATCTTGTGGATCGCGGCCATCCGCATCGTTTCCATGACGATCA	927
Db	122107	CAGCACGGGCGGTTGCAACCTCGCGAGGACTCCATCCGATTTGTTTCCCTCACGGTGA	122048
Qy	928	CGGAAGCGGCTACAAATCAACATCAACGACGACCGGTGCTTCGATCTGGAGATCGGCGAG	987
Db	122047	CTGAAGCGGATACAATCATGATCCGGCGACAGAAGATTTCGACACACCAACCCCTCGAA	121988
Qy	988	TAAAGGCGGACCTCAAGAACCGGAAAGCGGTCT-----ACCGTTTTTCG	1032
Db	121987	TCGTTGCTACCGCGAAGCCCTGCAGGGGCGATACCTCCACTTTCGAGACCTTCTTTG	121928
Qy	1033	GTTACGTGGTGGAGCCCTCGCTCGTCTGTGGGATGCCGTGGTAAGGCATTTACGTCFA	1092
Db	121927	GGTTGATCACTCGCGCATTGATTTCCCGAAAGAATCAGGATCTAGGCCATTTTACCATCA	121868
Qy	1093	TGTCCTGTGATAACCTGGTCAFAAAGGCAATGTGCCCGCAAGGCTTCTCTGGCTATG	1152
Db	121867	TGAGCTCGGATAACATCCAAGGCAACGGGATCTGGCTAAGCGTTTCTTCTCGCCCTTCG	121808
Qy	1153	CGAAGGCGCGGATCCGGAGTTGCGAAGTCGATTGAGGAAACCGGACCTTCCCGAAGC	1212
Db	121807	CACATTCGTGCTTCTTGAGCTCGGGAATGGGTGGAACAACGTGGCCTTCCCAACT	121748
Qy	1213	GAATGGTTGATCGATCACCCGACCGGTTTCGGCGGAATCGCCAAAGATCAACGCGG	1272
Db	121747	CCATGGTGACCGCATCACCC---CTGAACACACCGACGGCGCGCATCACATCAAGG	121691
Qy	1273	CCAGTGGGCTGGATGACGACCTCGCGCTGTGGCGGAGGATTTCCATCAGTGGGTGCTGG	1332
Db	121690	AAATCGGCTACATCGATCGCTGGCGAGTGTTCTGAAAGATTTTCAACCAATGGGTCTCG	121631
Qy	1333	AAGACCAAGTTTTCGGATGGCGGCTCCGCGCTTGAAAAAGCGCGGTGCAGATGGTTCGGG	1392
Db	121630	AGGATGCCTTTCAACCGAGGCGCGCCCGCGTACGAGGAGTTGGCGTGCAGTCTCTCCG	121571
Qy	1393	ACGTGACGAGCTGGGAGTACGTCAAGATCCGAATGCTCAATGCAAGGGCATGTCTGCTCT	1452
Db	121570	ACGTGGAGCCCTTATGAATTAATGAAGCTGCGCCCTGCTCAAGGCTTCCACACGAGCATTT	121511
Qy	1453	GCTTCCGAGCATTTCTGTCGGCTATGAAATGTGGATGACGCCATTCGAACACACGGAAC	1512
Db	121510	GCTACTTCGGGCATTTGCTGGCCACCATGTTCCACGAGCTATGGCGGATACCCGCT	121451
Qy	1513	TCCTTGGCAATCTGAAGAACTATCTCAACAAGAGTGTATCCCGACCCCTGAAGCGCCCT	1572
Db	121450	TCCAGGATTTCTCTCTGCTTACATGGAGCGGAGCCACCCCTACCTTCAAGGAATTC	121391
Qy	1573	CAGGATGACGCTCGAAGGCTATCGGAGACGGTTCATCAGCGGTTTCTTCAACACAGGGA	1632
Db	121390	CAGGTGTCGATCTAGATCTTATCGACGCCAACTCATCGCGGATTCGCGCAACGCGCGAG	121331
Qy	1633	TGTCGGACACAGCCTCCGGATTTGCTACGATGGCTGTTTCAAAGTTTCAAGTGTCTCGA	1692
Db	121330	TCAAAGACACCGTACCGGCGCTGTGTGGGAATCTCTCGACCGAATTCAAAGTGGCTGT	121271
Qy	1693	CGGAAACCGCTCGGCGGATCGAAGACAAGCGGGACCTGTACGATATAGGCTTCGGAA	1752
Db	121270	TGCCAGTCGTACGGGAARACCTCGACAGAGCGCGACGTACACTTCTTCGAGCCATCG	121211
Qy	1753	TTGCATCTATCTCGAATGCTGGGTGCGGACGAGAGGGC	1796
Db	121210	TCGATCTCTGGCGCGCTTACGACGAAGGACCGACGACGAGCGGC	121167

DE DNA encoding novel human diagnostic protein #28804.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

KW Homo sapiens.

OS

XX WO200175067-A2.

XX

XX 11-OCT-2001.

XX

XX 30-MAR-2001; 2001WO-US08631.

XX

XX 31-MAR-2000; 2000US-0540217.

XX

XX 23-AUG-2000; 2000US-0649167.

XX

XX (HYSE-) HYSEQ INC.

XX

XX Drmanac RT, Liu C, Tang YT;

XX

XX WPI: 2001-639362/73.

XX

XX P-PSDB; ABG28813.

XX

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX biodiversity

XX

XX Claim 1: SEQ ID NO 28804; 103pp; English.

PS

XX

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX (II). (II) is useful for generating antibodies against it, detecting or

XX quantitating a polypeptide in tissue, as molecular weight markers and as

XX a food supplement. (II) and its binding partners are useful in medical

XX imaging of sites expressing (II). (I) and (II) are useful for treating

XX disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits to assess biodiversity

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. AAS64197-AAS94564 represent novel human

XX diagnostic coding sequences of the invention.

XX Note: The sequence data for this patent did not appear in the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published pct sequences.

RESULT 8

AAS93000

AA353009
ID AAS93000 standard: cDNA: 3390 bp.

XX
ID
XXAA
AC

AL XX AAS

XX
DT 13-XX
D.F. 2007-2007 (TTSU enry)

XX

QY 799 CTCCTCGCGCAAGACACGCTGCGCTCATATGGGCGCGCTCGTACATATCTGTGCCC 858
 DB 2723 GTGCGGACGCAATCAGGTGATTAATTTGCGTGGCGTTTACGAAATGCTTAATGCAAAAC 2782
 QY 859 CGGCGGATCCGGAAGCGCTGCTGAGCATCTTGTGATCCGCCATCCGATCTTTCCA 918
 DB 2783 TGGATTCTTACGCGCAATTAATGAGAAATTTTGGAGCCACAGGTGGCAATTTGTTCCC 2842
 QY 919 TGACGATCACGAAGCGGCTTACACATCAACGAGACGACCGGTCCGTTTCATCTGGAGA 978
 DB 2843 TGACGATTACCGAAAAAGGCTATTGTAATGACCGCCACCGGTGCACTCGACACAGTA 2902
 QY 979 ATGCGGAGTAAGCGCGACCTCAAGAACCCGGAAGCCCTCTACCGTTTTCGGTTACG 1038
 DB 2903 ATCCGCGGATTAATTCACGATACAAACCCCTGAAGAACCTCACTCCGACCGGTATTC 2962
 QY 1039 TGGTCGAGGCGCTGCTGCTGTTGGGATCCGCTGAGGATTAAGGCATTTACGCTCATGCTCCT 1098
 DB 2963 TCGTGAAGCACTGAAGCCCGCGCTGAGCGCGCTTACACCGTTTACCGCTGCTCTCCT 3022
 QY 1099 GTGATAACCTCGCTATCAACGCAATGTGCGCCGCAAGCGCTTCTCGGCTATCGAAGG 1158
 DB 3023 CGGACAATATCCCGACAATGTCATGTTGTTGAAAGCGCGTCTGGGATGCGAGAA 3082
 QY 1159 CGCGGATCCGAGTTGGCGAAGTGGATTGAGGAAAGCGGACCTTCCCGAACGGAATGG 1218
 DB 3083 AGCGTTGCCAGAACTCGCCGGGTGGATAAAGACAGCAGTCAGTCTTCCGGGAACCATGG 3142
 QY 1219 TTGATCGCATCACCCGACCGCTTTTCGGCGAA 1250
 DB 3143 TCAGACCGCATTTGTTCCCGCTCAACCGACGAA 3174

RESULT 9

AA93884/C
 ID AA93884 standard; cDNA; 3390 BP.
 AC AA93884;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE DNA encoding novel human diagnostic protein #29688.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR P-PSDB; ABG29697.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX
 PS Claim 1; SEQ ID No 29688; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. AAS64197-AAS94564 represent novel human
 CC diagnostic coding sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 3390 BP; 792 A; 909 C; 925 G; 764 T; 0 other;

Query Match 2.9%; Score 118.4; DB 23; Length 3390;
 Best Local Similarity 50.4%; Pred. No. 1.6e-20;
 Matches 349; Conservative 0; Mismatches 331; Indels 12; Gaps 2;
 QY 562 CTCCTTCCTGCGCAATGTCAGGCTCCCGCTATGATCGACGGATCAAGCTTGGGATCG 621
 DB 899 CGCTCCCGCATCATGTACACGCTCCACGCTATGATCGCCACGATGGAATCATCATCG 840
 QY 622 TGCATTTTCGGTGTAGGTAACATTTTTCGAGCCCATGAGCGCTTCTACGTCGAGCAGATTC 681
 DB 839 TTCAATTTTGGCTTTGGAGCCCTTTCACCGCGCTCATCAGGCGTTACTGACCGCATCTG 780
 QY 682 T---TGAACACGCTCCGGATGGCGGATTTGTTGGTGTGGCTGACGGGGAGTGACCGTT 738
 DB 779 TGAATGCCAGGCTAGGCGACTGGGGGATCTCTGAAATCAGCTTGTTCAGCGGTGATCAAC 720
 QY 739 CAAAGAAAAAGCCGAGGAATTCAGAGGCCAGGACTGCTGTATTTCCCTACCGCAGAGCG 798
 DB 719 TGATGAGCCAGC-----TCCCGCGACAGAACCATTTATATACCGTGTGAGAAAG 669
 QY 799 CTCCGTCGCGCAAGACGACGCTGCGCTCATATGGGCGCGCTCGGTACTATCTGTGCCCC 858
 DB 668 GTGCGGACGCAATCAGGTGATAATTTGTCGGTGGCTTACAGAAATGCTTAATGCAAAAC 609
 QY 859 CGGCGGATCCGGAAGCGCTGCTGAGCATCTTGTGATCCGGCCATCCGCATCTGTTTCCA 918
 DB 608 TGGATTCTTACGCGCAATTAATGAGAAATTTTCCGAGCCACAGGTGGCAATTTGTTCCC 549
 QY 919 TGACGATCACGAAGCGGCTTACAAATCAACGAGACGACCGGTGCGTTTCATCTGGAGA 978
 DB 548 TGACGATTACGAAAAAGGCTATTGTTGACCGCGCCACCGGTGCACTCGACACAGTA 489
 QY 979 ATGCGGAGTAAGCGCGGACCTCAAGAACCCGGAAGCCCTCTACCGTTTTCGGTTACG 1038
 DB 488 ATCCGCGGATTAATTCACGATCTACAAACCCCTGAAGAACCTCACTCCGACCGGTATTC 429
 QY 1039 TGGTCGAGGCGCTGCTGCTGTTGGGATGCCGTTGGTGAAGGCATTTACGGTCAATGCTCCT 1098
 DB 428 TCGTGAAGCACTGAAACCGCGCTGAGCGCGCTTACACCGTTTACCGTGTCTCCT 369
 QY 1099 GTGATAACCTCGCTATCAACGCAATGTGCGCCCGCAAGCGCTTCTCGGCTATCGAAGG 1158
 DB 368 CGGACAATATCCCGACAATGTCATGTTGTAAGAACCGCGGTCTGGGATGCGAGAA 309
 QY 1159 CGCGGATCCGAGTTGGCGAAGTGGATTGAGGAAAGCGGACCTTCCCGAACGGAATGG 1218
 DB 308 AGCGTTCCCGAGAACTCGCGGGTGGATAAAGACAGCAGTCAGTCTTCCGGGAACCATGG 249
 QY 1219 TTGATCGCATCACCCGACCGCTTTTCGGCGAA 1250
 DB 3143 TCAGACCGCATTTGTTCCCGCTCAACCGACGAA 3174

Db 248 TCGACCGCATGTTCCGGCTGCAACCGACGAA 217

RESULT 10
AAD16810
XX AAD16810 standard; DNA; 1848 BP.
XX
XX AAD16810;
XX
XX 29-NOV-2001 (first entry)
XX
XX Escherichia coli strain C arabinol dehydrogenase gene.
XX
XX Positive selection system; metabolise; arabinol; ribitol; mannitol;
KW transgenic cell; marker gene; arabinol dehydrogenase; ds.
XX
XX Escherichia coli C.
XX
XX WO200166779-A2.
XX
XX 13-SEP-2001.
XX
XX 08-MAR-2001; 2001WO-US07474.
XX
XX 08-MAR-2000; 2000US-0188291.
XX 15-AUG-2000; 2000US-0255595.
XX
XX (UYGE-) UNIV GEORGIA RES FOUND INC.
XX
XX Parrott W, Lafayette P, Kane P;
XX
XX WPI; 2001-565596/63.
XX
XX Positively selecting transformed cells comprising selectable marker
PT gene and desired gene, from a cell population by using marker compounds
PT e.g., arabinol, ribitol which confer selective advantage on transformed
PT cells -
XX
XX Claim 1; Page 34-35; 37pp; English.
XX
XX The present invention relates to a positive selection system that
CC involves conferring to transformed cells the ability to metabolise
CC arabinol, ribitol and/or mannitol. The positive selection method is
CC used in positively selecting transgenic cells from a population of cells
CC using the positive selection method, the presence of the gene of
CC interest in the genetically transformed cells may be determined without
CC the disadvantages associated with traditional negative selection
CC systems. Positive selection of the transformed cells is achieved without
CC directly damaging the neighbouring non-transformed cells. The
CC transformed cells may be identified by simple visual means without the
CC use of a separate assay to determine the presence of a marker gene. This
CC technique also avoids the release of antibiotics or other dangerous
CC genes into the environment. The present sequence is Escherichia coli
CC strain C arabinol dehydrogenase gene.
XX
XX Sequence 1848 BP; 494 A; 443 C; 473 G; 438 T; 0 other;
SQ

Query Match 2.6%; Score 105; DB 22; Length 1848;
Best Local Similarity 46.8%; Pred. No. 4, 1e-17;
Matches 509; Conservative 0; Mismatches 560; Indels 18; Gaps 5;

QY 622 TCGATTCGGTGTAGTAACCTTTTCGAGCCCATGAGCGCTTACGTCGAGCAGATTC 681
DB 503 TGCACATCGGTTAGTGTCTTTTCATCGCGCACATCAGCGGTGGTATCTACACCGTTTGC 562

QY 682 TTGACACCGCTCCGACTCGGCGATGTTGGTGTGGCTGACGCGCAGTGCCTTCAA 741
DB 563 AGGTGATGGCGGATAAAGCTGGAGCATTTGCTGGCGCAATATCTGATATGA---TGCTG 619

QY 742 AGAAAAAGCCGAGGAATTCAGAGCCAGGACTGCTGTATTTCCCTGACCGAGCGGCTC 801
DB 620 ACATGTCTGACAGGCATCAGTCGACACAAAGGTCGTATGTGCTGGAAACCGTCAGCC 679

QY 802 CGTCCGGCAAGACACGCTGCGCGCTCATGGCGCGCTGCGTGACTATCTGCTTCCCGCG 861
DB CGGAAGGGTAAAGCAATATGAGAGATCACTCAATTGAGAGTTGATACCGTGGCAGG 739

QY 862 CCGATCCGGAAGCGCTGCTGAGCATCTTGTGATCCGGCCATCCGATCTGTTCCATGA 921
DB 740 CAGATTTACACCGCTGATTGCTGAAGGGCAGATCCGAGACAAAGTGAATGCTTTCA 799

QY 922 CGATCAGGAAGGCGCTTACAACATCAACGAGACGCGGTGCGTTGATCTGAGAGATG 981
DB 800 CCGTACCGAAGCGGGTACTACCTGA---ATACAGTACAAACTGAAGTAAACAATC 856

QY 982 CGGAGTAAAGCGCGACTCAAGAACCCGGGAAAGCGCTCTACCGTTCCTGGTTACGTGG 1041
DB 857 CTGATTTAGCGGAGATCTTTAAGGGGATGCAAAACAATTTACGGTGTATTACCCGTA 916

QY 1042 TCGAGGCCCTGCTGCTGCTGTTGGGATCGCGGTGTAAGGCAATTTACGGTCTATGCTCTG 1101
DB 917 TCCTCGAAGCGGTATGGCAAAATACGCCGG-----ACCACTAACCTGCTGAATGCG 970

QY 1102 ATAACTCGCTCATAAGCGCAATGTGCGCCCAAGGCTTCTCTCGGCTATCGAAG---G 1158
DB 971 ATACGTGCGCCATAATGTTGACGTTTCCATGATGGCTGCTGTTGAGTTTCTCCAGCTAA 1030

QY 1159 CGCGCATCCGAGTTCGCGAAGTGGATTGAGGAAACCGGACCTTCCCGAACGGAATGG 1218
DB 1031 CTGGCAACAGGATGTCATCGACTGGCTGAGTACAAATACCAATTTGCCCGCAATACCATGG 1090

QY 1219 TTGATCGCATCACCCCGACCGCTTTTCGGCGGAAATCGCCAAAGCTCAACGCGCAGTG 1278
DB 1091 TTGACCGCATTACCGCTGCTCGGACGAGAACTTCCGGCAGGATCAAGGCTCAACGCG 1150

QY 1279 GGCTGGATGACAGCTGCGCTGGTGGCCGAGGATTTCCATCAGTGGGTGCTGAAGACC 1338
DB 1151 GTATTGCGATAAAGCGCGGTAAATGGCGCAACCTTTATCCAGTGGTCTGTTGAAGATA 1210

QY 1339 AGTTTGGGATGGCGCTCCGCGCTTGAAAGCCGCGCTGCAGATGCTCGGGACGTGA 1398
DB 1211 ATTTCCGTGATGTCGCTCCGCACTGGAGAGGTCGGTGTGAACTGGTGGCTCGGTAA 1270

QY 1399 CGGACTGGGAGTACGTCGAATGCCAATGCTCAATGACGGCATGCTATGCTGCTTCC 1458
DB 1271 TCCCTATGAGAGCGGGAAGATTCGATTTTAACTCTTACACAGTTGCTGCTGGG 1330

QY 1459 CAGGATTTCTGGTGGCTATGAGATGAGATGGATGAGCCATTGAGACAGCGAATCTCTG 1518
DB 1331 CAGGTACGTTAATCGGTCAAAATATATCCAGAAAGCAATGACCGATTTTATCTATC 1390

QY 1519 GCAATCTCAAGAACTATCTCAACAGGATGTCATCCCGACCTGAAGCGCTTCCAGCA 1578
DB 1391 AGATTGGCGCGCTTACGTGACGGAAGATGTCATTTCTTGGTT---GGCGATACGGTA 1447

QY 1579 TGACGCTCGAAGGCTATCGGACAGCGCTCATCAGCGTTTCTTCCAAAGCGGATGTCGG 1638
DB 1448 TCGATTTGCCAACCTACCGTGTGTTGTTACTCAACCGTTTTTACCAATCCACATATTCAGG 1507

QY 1639 ACCAGACGCTCGGATTCGTAGCGTGGCTGTTTCCAGGTTTCAGGTGTTCTGCGAGGAA 1698
DB 1508 ACACCAACCAACGCGCTCGGTGCGGATGGTTTCTCGAAAAATTCGCGGATGATTGCGCCCA 1567

QY 1699 CCGTGG 1705
DB 1568 CACTGCG 1574

RESULT 11
AAF12753
ID AAF12753 standard; cDNA; 684 BP.
XX
XX AAF12753;
XX AC
XX DT 13-MAR-2001 (first entry)
XX

17 GAACCTTTCCGCGCTGGTATAGACATAACTTTGTCGCGAGCTCCGGAATGGAA 76
1368 AAAGCCGCTGCAGATGGTGGGAGCTGAGGAGTACGTCAGATCCGATG 1427
77 AAAGCGGAGCGGAACTGGTGGGAGTGGTGGCTTATGAAGAGATGAAGTGGCGATG 136
1428 CTCATGAGGCGATGTCATGCTGCTTCCAGGCACTTCTGGTGGGCTATGAGAATGTG 1487
137 CTCACGCGCATTCATCTCTGCGGTATCTGGGTATCTTGCAGGATATCAGCATTT 196
1488 GATGAGCGCATTTGAAGACAGCAACTCTCTGGCAATCTGAAGAACTATCTCAACAAGGAT 1547
197 AATGACTGTATGAAGAGATGAACATATCTTATGCGGCGTATGGCTTGATCTTCAGGAA 256
1548 GTCATCCGCGCTGAGGCGGCTTCAGGCGATGAGGCTGAGGCTATCGGAGCAGCGTC 1607
257 CAAGCGCGGAGCTTGAAGAGTGC---AGGGCGTTGATTTGCAAGATTACGCTTAACCGATT 313
1608 ATCAGCGCTTTCTCAACAAGCGGATGTCGACGAGCGCTCCGATGCTAGCGATGGC 1667
314 ATTGACGCTATAGCAACCGCGGTAGCTCATCGAACCTGGCAGATTGCGATGGATGGT 373
1668 TGTTCAGGCTAGGCTTCTGACGAGAAACCGTGGCTGCGGCGATCGAAGAACGCGG 1727
374 AGCCAGAAATGCCACAGCGGATGTTGATTTCTGTCGCTGGCATCTGGCGCATGACAGC 433
1728 GACCTGTACGTATAGCGTTCGGAATTCGATCTCATCTCGAAATGCTCGGTGGCGAG 1787
434 AAGTTCGATCTGTCGCGCTGGCGTGGCGTGGATGCGTTATGTCGGTGGTGGAT 493
1788 GAGAGGCGGCGACGATGAA 1808
494 GAACAGGAAATCCGATGAA 514

RESULT 13

AAAF08165
ID AAF08165 standard; cDNA: 621 BP.
XX
AC AAF08165;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:688.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
XX
PI Berka RW, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI: 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX

PS
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Trichoderma reesei; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 621 BP; 152 A; 199 C; 123 G; 142 T; 5 other;

Query Match 1.5%; Score 62.8; DB 21; Length 621;
Best Local Similarity 49.4%; Pred. No. 3.2e-06;
Matches 213; Conservative 0; Mismatches 212; Indels 6; Gaps 2;
Oy 576 GTCCAGGCTCCCGCTATGACATCGACGGGATCAAGCTGGATCGTTCGTTGTA 635
197 GTCCAGGCTCCCGCTATGACATCGACGGGATCAAGCTGGATCGTTCGTTGTA 256
636 GGTAACTTTTTCGAGCCCATGAGGCGTTCTACGTGAGCAGATTTCTTGAACAGCTCCG 695
257 GCGGTTTCCACAGAGCTCACTTGGCTCTACGTGCGACAAGCTGCTTGAGAAGCAGCGT 316
696 GACTGGCGGATTTGTTGGTGGCTGACGGGCGATGACCGTTCAAGAAAAACCGAG 755
317 GAGCGGATTTGGGCGCATCTGCGGTATGCGCTCCGCTCCCAACGATGTGCGCATCGCGAT 376
756 GAATTCAGGCGCCAGGACTGCTGTATTCCCTGACCGAGCGGCTCCGTCGCGCAAGAGC 815
377 GTTCTCAACGACAGGACCACTCTACACCGTATTGAGCGATCCCNAAAGGTAGCTTC 436
816 ACGGTGCGGCTATGGGCGCGCTGCGTGACTATCTGTTGCCCGCGCGATCCGGAAGCG 875
437 GCGGATGTCGTC---GGCAGCATCACTCTTCTTTCGCGCTGATGACCGTGAGGCT 492
876 GTGCTGAAGCATCTGTTGATCCGCGCATCGCATCTTTCATGACGATCAGGAAGGC 935
493 GGCATTGCAAGATGGGNTCATCGACACTACATTTGTTTTT--TTACTATTACTGAGAGN 550
936 GCGTACAACATCAACGAGACGCGGTGCGTTGCTGATCTGGAAGATGCGCGCAGTAAAGGCC 995
551 GGTACTACTACAGGAGAACACCCATCAATTTGGAAGCTGAGCATTTCTGGCATTCAGCAC 610
996 GACCTCAAGAA 1006
611 GACTTCAAGAA 621

RESULT 14

AAS92759/c
ID AAS92759 standard; cDNA: 921 BP.
XX
AC AAS92759;
XX
DT 13-FEB-2002 (first entry)

XX	DNA encoding novel human diagnostic protein #28563.
DE	
XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX	
XX	Homo sapiens.
XX	
PN	W0200175067-A2.
XX	
PD	11-OCT-2001.
XX	
XX	30-MAR-2001; 2001WO-US08631.
PF	
XX	31-MAR-2000; 2000US-0540217.
PR	
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Drmanac RT, Liu C, Tang YT;
XX	
XX	WPI: 2001-639362/73.
DR	P-PSDB: ABG28572.
DR	
XX	
XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity
XX	
XX	Claim 1; SEQ ID No 28563; 103pp; English.
PS	
XX	
XX	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and gene mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. AAS84197-AAS94564 represent novel human
CC	diagnostic coding sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
XX	Sequence 921 BP; 230 A; 235 C; 230 G; 226 T; 0 other;
SO	
Query Match 1.3%; Score 53.8; DB 23; Length 921;	
Best Local Similarity 53.6%; Pred. No. 0.0009;	
Matches 112; Conservative 0; Mismatches 97; Indels 0; Gaps 0;	
QY	1307 CGAGGATTTCCATCAGTGGGTGCTGGGAAGACCAGTTTGGCGATGGCCGTCGCGCTTGA 1366
Db	907 CGAACCGTTTATCCAGTGGGTGGTGGGAAGATAACTTCGTCGCTGGGCGTCTGCGCTGGGA 848
QY	1367 AAAAGCGCGCTGCAGATGCTCGGGGACGTGACGGACTGGGAGTACCTCAAGATCCGAAT 1426
Db	847 AGTCGCAGGTGTACAATGTGATGATGCTTCGCCATGGGAGAGATGAACATGCCGAT 788
QY	1427 GCTCAATGCAGGCGATGTCATGCTCTGCTTCCAGGCAATTCGTGGCTGATGAGAAATGT 1486
Db	787 GCTTAATGGCAGCCACTCTTTTTCGCTTATCTGGGTTACTCTCAGATTCGCCCATAT 728
QY	1487 GGATGACGCCATGGAAGACAGCAGCAATCC 1515

```
QY 1382 GATGTCGGGACGTGACGGACTGGGAGTACGTCAAGATCGAATGTCTCAATGCAGGGCA 1441
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
Db 379 AATGCTGAATGATGCTCTGCCATGGAGAGATGAAGATGCGGATGCTTAATGGCAGCCA 320
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
QY 1442 TGTCAATGCTCTGCTCCAGGCATTCCTGGTCGGCTATGAGAAATGTGGATGACGCCATTGA 1501
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
Db 319 CTCCTTTCTCGCTTATCTGGGTTACCTCTCAGGATTCGCCCATATCAGTGATTGCATGCA 260
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
QY 1502 AGACAGCGAATCCTTGGCAATCTGAAGAACTATCTCAACAAGGATGTCATCCCGACCCCT 1561
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
Db 259 GGATCGCCGCA---TTTCGCCATGCCGCCAGAACATTAATGTGGATGAGCAAGCGCCGAC 203
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
QY 1562 GAAGGCGCTTCAGGCATGACGCTCGAAGGCTATCGGGACAGCGTCAATCAGCCGTTTCTC 1621
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
Db 202 ACTGCAAAATTAAGATGTCGATTTAACACAAATATCCGGATTAAGTTAATTGCACGTTTGC 143
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
QY 1622 CAACAGCGGATGTCGGACGACAGCGCTCCGGATTGCTAGCGATGGCTGTTCCAAAGGTTCA 1681
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
Db 142 TAATCCGCGCTGAAACATAAGACCTGGCAAAATCCGATGGATGGCAGCCAGAAATTACC 83
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
QY 1682 GGTGTTCTGGACGGAACCGTCGTCGGGCGATCGAAGACAAGCGGGACCTGTCAAGTAT 1741
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
Db 82 GCACCGCATGCTGGCAGGTATTCGCATACATCAAGGGCGCCGAACGGACTGTCGTTGCT 23
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
QY 1742 AGCGTTCGGAATTCATCCTA 1762
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
Db 22 GGCATTACGAGTCACGTCTTA 2
    |||||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||  ||  |||||  ||  ||
```

Search completed: May 18, 2003, 23:47:14
Job time : 1197 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 16:35:29 ; Search time 177 seconds
(without alignments)
7129.802 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttcagtcctgcaggt.....tcggggcggtcctgaagctt 4115

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	44	1.1	1926	4	US-09-249-585A-4
C 2	44	1.1	1931	2	US-09-130-114-2
C 3	43	1.0	477	4	US-09-135-994-1
C 4	42.6	1.0	2159	3	US-08-286-870A-7
C 5	41.2	1.0	533	6	5482709-5
C 6	40.4	1.0	543	6	5273901-6
C 7	40	1.0	234	1	US-08-469-802B-3
C 8	40	1.0	234	2	US-08-267-803B-3
C 9	39.6	1.0	4403765	4	US-09-103-840A-2
C 10	39.6	1.0	4411529	4	US-09-103-840A-1
C 11	39.4	1.0	298	2	US-08-637-759B-21
C 12	39.4	1.0	298	3	US-08-871-355A-21
C 13	39.4	1.0	298	4	US-09-201-945-21
C 14	38.8	0.9	397	3	US-09-232-691-3
C 15	38.6	0.9	1794	4	US-09-232-191-36
C 16	38.6	0.9	1794	4	US-09-232-200-90
C 17	38.6	0.9	1794	4	US-09-232-197-90
C 18	38.6	0.9	1794	4	US-09-232-201-90
C 19	38.6	0.9	2007	4	US-09-232-191-22
C 20	38.6	0.9	2007	4	US-09-232-200-22
C 21	38.6	0.9	2007	4	US-09-232-197-22
C 22	38.6	0.9	2007	4	US-09-232-201-22
C 23	38.6	0.9	8540	4	US-08-487-283A-4
C 24	38.6	0.9	8540	5	PCT-US96-05611A-12
C 25	38.6	0.9	8932	2	US-08-252-493C-8
C 26	38.6	0.9	8932	3	US-09-276-197-8
C 27	38.6	0.9	4403765	4	US-09-103-840A-2

28	38.6	0.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
29	38.4	0.9	2870	5	PCT-US93-07213-3	Sequence 3, Appli
30	38.2	0.9	2799	4	US-09-232-279-1	Sequence 1, Appli
31	38.2	0.9	3382	2	US-08-682-847-1	Sequence 1, Appli
C 32	38	0.9	3010	4	US-09-199-637A-168	Sequence 168, App
33	37.8	0.9	279	4	US-09-481-288-14	Sequence 14, Appl
34	37.8	0.9	932	1	US-08-114-692A-3	Sequence 3, Appli
35	37.8	0.9	932	2	US-08-723-306-3	Sequence 3, Appli
36	37.8	0.9	932	5	PCT-US96-10041-3	Sequence 3, Appli
C 37	37.8	0.9	1693	3	US-09-320-878-23	Sequence 23, Appli
38	37.8	0.9	30001	1	US-08-125-468-1	Sequence 1, Appli
39	37.8	0.9	30001	2	US-08-474-933-1	Sequence 1, Appli
C 40	37.6	0.9	3366	1	US-08-469-802B-1	Sequence 1, Appli
C 41	37.6	0.9	3366	2	US-08-267-803B-1	Sequence 1, Appli
C 42	37.6	0.9	10348	2	US-08-457-273B-41	Sequence 41, Appli
C 43	37.6	0.9	10348	3	US-08-556-419-13	Sequence 13, Appli
C 44	37.6	0.9	10348	4	US-09-041-886-14	Sequence 14, Appli
C 45	37.6	0.9	10366	1	US-08-246-982A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: Template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match	1.18;	Score 44;	DB 4;	Length 1926;
Best Local Similarity	48.18;	Pred. No. 0.031;		
Matches 125;	Conservative 0;	Mismatches 135;	Indels 0;	Gaps 0;
Qy	1154	GAAGCGCGCGATCCGGAGTTGGCGAAGTGTGAGGAAACGCGACCTTCCCGAACGG	1213	
Db	1022	GGAGGAGCTGGGCGCGGAGGTGACGGAGGAGTGGGCGCGGAGGTGGAGGACGAGGACGG	963	
Qy	1214	AATGTTGATGCGATACCCCGACCGTTCGGCGGAATGCCAAGAGCTCAACGCGGC	1273	
Db	962	GGAGGACGACGACGCGGGAGGAGGAGGACGCGGGAGGACGCGGGAGGACGAGGAGGGA	903	
Qy	1274	CAGTGGCTGGATCAGCACCTGCCCTCGTGGCGCGAGGATTTCCATCAGTGGGTCTGGA	1333	
Db	902	GGACGGGAGGACGAGGACGCGGGAGGAGGAGGACGCGGGAGGACGCGGGAGGAGGA	843	
Qy	1334	AGACGAGTTTGGCGATGCCCTCCCGCTTTGAAAAACCGCGCTGCAGATGATGTCGGGGA	1393	
Db	842	GGACGAGGACGCGGGAGGAGGAGGAGGAGGACGCGGGAGGACGCGGGAGGACGAGGA	783	
Qy	1394	CGTGACGACTGGGAGTACG	1413	
Db	782	CGGGAGGACGCGGGAGGACG	763	

RESULT 2
US-09-130-114-2/c
; Sequence 2, Application US/09130114

[illegible]

```

; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2159
; US-08-286-870A-7

Query Match 1.08; Score 42.6; DB 3; Length 2159;
Best Local Similarity 54.0%; Pred. No. 0.081;
Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

```


ORGANISM: Partial sequence of *Salmonella typhimurium*

QY 61 GATGGCGGCTGTGTTGACCGACATGATGCTGTGTGCGACGTGCGCATTTGGACGG 113
Db 72 GCCAGATCTGATCAAGACAGACATCCAACTCGGGCGGATCGCCATACGCCAG 124

RESULT 13
US-09-201-945-21
; Sequence 21, Application US/09201945
; Patent No. 6342215
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/201,945
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/637,759
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 298 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Partial sequence of Salmonella typhimurium
; ORGANISM: virulence gene
US-09-201-945-21

Query Match
Best Local Similarity 1.0%; Score 39.4; DB 4; Length 298;
Matches 67; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1 AGCTTGATGCTGCTCAGGTGACTCTAGAGATCCGGTTTTTGGCAGCGCTCCCTAGATT 60
Db 12 AGCTTGATGCTGCTCAGGTGACTCTAGAGATCCGGTTTTTGGCAGCGCTCCCTAGCG 71

QY 61 GATGGCGGCTGTGTTGACCGACATGATGCTGTGTGCGACGTGCGCATTTGGACGG 113
Db 72 GCCAGATCTGATCAAGACAGACATCCAACTCGGGCGGATCGCCATACGCCAG 124

RESULT 14
US-09-253-691-3/c
; Sequence 3, Application US/09253691
; Patent No. 6124100

; GENERAL INFORMATION:
; APPLICANT: Dong Kyu JIN
; TITLE OF INVENTION: Diagnostic Method and Kit for Neuropsychiatric Diseases
; TITLE OF INVENTION: Using Trinucleotide Repeats Sequence
; FILE REFERENCE: 1942/36
; CURRENT APPLICATION NUMBER: US/09/253,691
; CURRENT FILING DATE: 1999-02-22
; EARLIER APPLICATION NUMBER: KR 98-6,278
; EARLIER FILING DATE: 1996-02-26
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: WordPerfect 6.1/Windows
; SEQ ID NO 3
; LENGTH: 397
; TYPE: DNA
; ORGANISM: human
US-09-253-691-3

Query Match 0.9%; Score 38.8; DB 3; Length 397;
Best Local Similarity 48.2%; Pred. No. 0.38;
Matches 109; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 3337 CTGATGGATCTTTGATCAAGCGTCTGCCAATGTGGCTGCAGAAAGTCTGAACCTGGTTG 3396
Db 349 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 290
QY 3397 CGGAAGCGGATCAATAATGGTTCCGGATTCGGGCGGCGTGTCTTCATGCTGGCGGC 3456
Db 289 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 230
QY 3457 GTTCTGTCCATCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3516
Db 229 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 170
QY 3517 GCGCAGGATATTCGGTTCTTCCTCCGCTTCAGGGCGCCCTCTTGC 3562
Db 169 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 124

RESULT 15
US-09-232-191-36
; Sequence 36, Application US/09232191
; Patent No. 6284487
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; TITLE OF INVENTION: Fatty Acid Transport Proteins
; FILE REFERENCE: WHI97-21p3ME
; CURRENT APPLICATION NUMBER: US/09/232,191
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 1794
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
US-09-232-191-36

Query Match 0.9%; Score 38.6; DB 4; Length 1794;
Best Local Similarity 50.3%; Pred. No. 0.95;
Matches 95; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 1139 CTTCTCTCGGCTATGCGAAGCGCGGATCCCGAGTTGGCGAGTTGGCGAGTTGAGAAACGC 1198
Db 1578 CTTGCCCGGCTATGCACATCCCGCTCTTTGTTGCGGTAGTGGGTGCTGCTGGCGCACACCAC 1637
QY 1199 GACCTTCCCGAAGCGAATGGTTGATCGCATCACCCCGACCCGCTTTTCGGCGGAATCGCCAA 1258

```

    ||| ||| || | | | | | | | | | | | | | | | |
1638 GACGTTCAAGAGTCGCAAGGTGGAGTTGCGCAACCAGGCCTATGGCGCGACATCGAGGA 1697
QY 1259 GAAGCTCAACGCGGCCAGTGGGCTGGATGACGACCTGCCGCTGGTGGCGCGAGGATTCCA 1318
Db 1698 TCCGCTGTACGTACTGGCGCGGCCGACGAGGATATGTGCCGTACTAGCGCGAATACCC 1757
QY 1319 TCAGTGGGT 1327
    | | | |
Db 1758 TGAGGAGGT 1766
```

Search completed: May 18, 2003, 22:35:23
Job time : 7951 secs

THIS PAGE BLANK (USF10)

GenCore version 5.1.6.
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 17:05:15 ; Search time 487 Seconds
(without alignments)
10899.580 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttgatgcctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 810007 seqs, 64496091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA: *

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	213.6	5.2	1509	9	US-09-738-626-126
2	213.6	5.2	3309400	9	US-09-738-626-1
3	105	2.6	1848	9	US-09-802-208B-1
4	42	1.0	2371	10	US-09-789-561-27
5	41.6	1.0	2614	9	US-09-822-846-491
6	41.4	1.0	1512	10	US-09-822-849A-393
7	41.4	1.0	1575	10	US-09-822-849A-261
8	41	1.0	501	10	US-09-833-790-158
9	40.4	1.0	1432	10	US-09-974-300-718
10	40.4	1.0	8017	9	US-10-152-040-26
11	40.4	1.0	8092	9	US-10-152-040-27
12	40.4	1.0	10417	9	US-10-152-040-28
13	40.4	1.0	536165	9	US-09-939-964-1
14	40.2	1.0	88421	9	US-09-976-059-1
15	40	1.0	3115	9	US-09-981-353-85
16	40	1.0	6604	10	US-09-880-107-1748
17	39.6	1.0	1146	9	US-09-738-626-2512
18	39.6	1.0	1854	9	US-09-894-844-103
19	39.6	1.0	3309400	9	US-09-738-626-1

20	38.8	0.9	724	9	US-10-123-155-60
21	38.6	0.9	1794	10	US-09-943-671-36
22	38.6	0.9	2007	10	US-09-943-671-22
23	38.4	0.9	2283	9	US-09-712-363-135
24	38.2	0.9	440	9	US-10-184-644-202
25	38.2	0.9	440	9	US-10-184-634-202
26	38.2	0.9	2799	9	US-10-085-519-1
27	38.2	0.9	5607	9	US-10-108-605-92
28	38	0.9	1049	9	US-10-123-155-358
29	38	0.9	3010	9	US-09-975-719-168
30	37.4	0.9	353	9	US-09-910-009A-157
31	37.4	0.9	694	9	US-10-184-644-60
32	37.4	0.9	694	9	US-10-184-634-60
33	37.2	0.9	415	10	US-09-960-352-1970
34	37.2	0.9	507	10	US-09-894-882-8
35	37.2	0.9	1503841	9	US-09-946-807-1
36	37.2	0.9	1503841	10	US-09-795-668-1
37	37.2	0.9	1503841	10	US-09-795-668-1
38	37	0.9	750	9	US-10-184-644-104
39	37	0.9	750	9	US-10-184-634-104
40	37	0.9	3796	9	US-10-127-391-32
41	36.8	0.9	18272	9	US-09-764-891-8869
42	36.6	0.9	941	9	US-10-123-155-464
43	36.6	0.9	3653	9	US-10-155-649-1
44	36.6	0.9	13910	9	US-09-919-901-1
45	36.6	0.9	13910	9	US-09-919-901-8

ALIGNMENTS

RESULT 1

US-09-738-626-126
Sequence 126, Application US/09738626
Publication No. US20020197605A1

GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO

APPLICANT: SENO, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09738.626

CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 126

LENGTH: 1509

TYPE: DNA

ORGANISM: Corynebacterium glutamicum

US-09-738-626-126

Query Match 5.2%; Score 213.6; DB 9; Length 1509;

Best Local Similarity 49.6%; Pred. No. 4.2e-56;

Matches 617; Conservative 0; Mismatches 609; Indels 18; Gaps 2;

QY 568 CTGCCAATGTCAGGCTCCCTTATGACATCGACGGGATCAAGCCTGGGATCGTGATT 627

Db 53 CTTCGGGAGTGCAGATCCCGGCTGACGCCGCTGACGTCGCCCGGGGATGTCCTACT 112

QY 628 TCGGTGTAGTAACTTTTTCGAGCCCATGAGGCGTTCTACGTCGACAGATTCTTGAAC 687

Db 113 TCAGTGTGGGGATTCATCGCGCTACCAAGGATGTACCTCAATGAATTGATG 172
Qy 688 ACGTCCGGAGTGGCGATTGTTGGTGTGGCTGACGGGAGTACCGTTCAAGAAA 747
Db 173 AGGCAAGGCGCTTGATTTGGGCGCATCATCGGCGATGGGTGTCATGCCCTCCGATGCGCA 232
Qy 748 AAGCGAGGAATCAAGGCCAGGACTGCTGTATTCCTGACCGAGAGCGCTCCGTCG 807
Db 233 TCGCGATGCGCTGGCGAGCAAGATCACCTTTATACCTGACCACTAAAGCTCCTGATG 292
Qy 808 GCAAGAGACAGGTGGCGGTATGGCGCGCTGCGTACTATCTGTTGCCCGCGCGATC 867
Db 293 GAATCTTTGATCAAAAAATCATCGATCCATCATTTGACTAGCTGTGCTCCCGAGGAC 352
Qy 868 CGGAAGCGGTGCTGAAGCATCTTTGATCGCGGCATCCGCATGTTTCCATGACGATCA 927
Db 353 CAGCAGCGCGCTTGCACCCCTCGCGAGGACTCCATCCGATTTGTTCCCTCAGGTGA 412
Qy 928 CGGAAGCGGTGCTGAAGCATCTTTGATCGCGGCATCCGCATGTTTCCATGACGATCA 987
Db 413 CTGAAGCGGTATCAACATCATCGATCCGCGACAGAGATTTGCGACACACCAACCTCGAA 472
Qy 988 TAAAGCGCGACTCAAGAACCGGAAAAGCGGTCT-----ACCGTTTTCG 1032
Db 473 TCGTTGCTGACCGGGAAGCCCTGACGGCGGCGATACTTCCACTTTGCGAGACCTTCTTG 532
Qy 1033 GTTACGTGGTGGAGCGCTCGCTGCTGTTGGGATCGCGTGGTAAAGCATTTACGGTCA 1092
Db 533 GGTGATCAGTGGCGCATTTGATTTCCCGAAAAGATCAGGATCTACGCCATTTACCATCA 592
Qy 1093 TGTCTGTGATAACCTGCGTCAATACGCAATGTCGCCGAAAGCGCTTCTCGCTCATG 1152
Db 593 TGAGTGCATACATCAAGCAACGCGGATCTGCTAAGGTTTCTTCCTCGCTCG 652
Qy 1153 CGAAGCGCGGATCGCGAGTTGGCGAAGTGGATTGAGGAAACGCGACCTTCCCGAAG 1212
Db 653 CACATTCGCTGTCTCTGAGTGGCGAATGGTGGAAAACAAAGTGGCTTCCCAACT 712
Qy 1213 GAATGTTGATCGCATACCGCGACCTTTCGGCGGAATGCCAAGAGCTCAACGGG 1272
Db 713 CCATGTGGAGCGGATCAACCC---CTGAACACCGGCGGCGGCGGATGACATCAAGG 769
Qy 1273 CCAGTGGGCTGATGACGACCTGCGCGTGTGGCGAGGATTTCCATCAGTGGGTGCTGG 1332
Db 770 AAATCGGCTACATCGATCGTGGCGAGTGGTTCTGAAAGATTTCACCAATGGGTCCCTCG 829
Qy 1333 AAGACAGTTTGGGATGGCGTCCGCGCTTGAAGAAACGCGGTGAGATGGTCCGGG 1392
Db 830 AGGATGCTTCAACCGAGGCGCGCGGTACGAGAGGTTGGCGTGCAGTCTCTCCG 889
Qy 1393 ACGTCAAGGACTGGGAGTACGTCAGATCCGAATGCTCAATGACGGGATGTCATGCTCT 1452
Db 890 ACGTGGAGCTTATGAATTAATGAAGTGGCGCTGCTCAAGCCCTCCACAGGAGCTTT 949
Qy 1453 GCTTCCCGAGGATTCGTGCGCTATGAGAATGGATGAGCCATTAAGAACAGCGAAC 1512
Db 950 GCTACTTCCGGCACTTGGTGGCGCAACACATGTTCCAGACGCTCATGCGGATACCCGCT 1009
Qy 1513 TCCTTGGCAATCTGAAGACTATCTCAACAGGATGTCATCCCGACCTGAAGCGGCTT 1572
Db 1010 TCCAGGATTTCTCTGCTTACATGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1069
Qy 1573 CAGGATGAGCTGCAAGGCTATCGGAGACGCTCATGAGGCTTCTTCCCAAGAGGGA 1632
Db 1070 CAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1129
Qy 1633 TGTGGAGACAGCGTCCGGATGCTAGCGATGGCTGTTTCCAAAGGTTTCAAGTGTCTGGA 1692
Db 1130 TCAAGACACACCGTACCGCGCTGTGTGCGGAATCTCCGACGCGATTTCCAAAGTGGCTGT 1189
Qy 1693 CGGAACACGCTGGTGGCGGATCGGAAGACAGCGGAGCTGTCAGTATACGGTTCGGA 1752

Db 1190 TGCAGTCTGTACGCGAAGAACTTCGACGAGCGCCGAGCTCACACTTTCTGACCCATCG 1249
Qy 1753 TTGCATCTATCTCGAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796
Db 1250 TCGCATCTTGGCGCGCTACGCAAGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1293
RESULT 2
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES.
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1
Query Match 5.2%; Score 213.6; DB 9; Length 3309400;
Best Local Similarity 49.6%; Pred. No. 1.1e-54;
Matches 617; Conservative 0; Mismatches 609; Indels 18; Gaps 2;
Qy 568 CTGCGAATGTCAGGCTCCCGCTATGACATCGACGGGATCAAGCTGGGATCGTGCATT 627
Db 122407 CTTCGGAGTCAGATCCAGCGTTCACACGCGTACAGCTGCGCCGGGCACTGTCCACT 122348
Qy 628 TCGGTGTAGTAACTTTTTCAGGCCATGAGGCGTTCAGCTGAGCAGATTTCTGAAC 687
Db 122347 TCGGTGTGGCGGATTCATCGCTCACCAAGGATGTACTCAATGAATTTGATGAATG 122288
Qy 688 ACGTCCGAGCTGGCGATTTGTTGGCTTGACGGGATGACCGTTCACAAAGAAA 747
Db 122287 AGGCAAGGCTTGGATTGGGCGATCATCGGATGGGTGTATGCTTCCGATGTGCGCA 122228
Qy 748 AAGCCGAGGAATCAAGGCCCGAGGACTGCTGTATTCCTTCCCGAGAGCGCTCCGTCG 807
Db 122227 TGGCGATGCTGGCGAGCCCAAGATCACCTTTATACCTTACCCTAACGCTCTCTGATG 122168
Qy 808 GCAAGAGCAGGTGGCGTCATGAGGCGCTGCGGTGACTATCTGCTGCCCGGCGGATC 867
Db 122167 GAATCTTTGATCAAAAATATCATCGGATCCATTTGACTAGCTGTTCCTCCCGAGGACC 122108
Qy 868 CGGAAGCGGTGCTGAAGCATCTTTGATCGGCGCATCCGCTGCTTTTCCATGACGATCA 927
Db 122107 CAGCAGGCGGTGCAACCTTCGCGGAGGACTCCATCCGATGTTTCCCTCAGGCTGA 122048
Qy 928 CGGAAGCGGTGCAACATCAACAGAGACCGGCTGCTGATCTCGAGATCGGAG 987
Db 122047 CTGAAGCGGATCAACATCATGATCGGCGGACAGAGATTTCCGACACCAACCCCTCGAA 121988
Qy 988 TAAAGCGCGCTCAAGAACCCGGAAAAGCGGTCT-----ACCGTTTTCG 1032

Db 121987 TCGTTGCTGACCGGAGCCCTGACGGCGGCGATACCTTCCACTTTGTCAGACCTTCTTTG 121928
QY 1033 GTTACGTGGTTCGAGGCGCTCGCTCGTGTGGGTGCGGTGATTAAGGCAATTTACGGTCA 1092
Db 121927 GGTGATCAGTCCGCAATGATTTCCCGAAGAGATCAGGATCTACGCCATTTACCATCA 121868
QY 1093 TGCTCTGTGATAACCTGCGTATACGGAATGTGCGCGGCAAGGCCCTTCTCGGCTATG 1152
Db 121867 TGAGTTCGATCAATCAATCAAGCAACGCGATCTGGCTAAGCGTTTCTCTCGCTTCG 121808
QY 1153 CCAAGCGCGCGATCCGAGTTGGGAGTGTGAGGAAACGCGACCTTCCGGAACG 1212
Db 121807 CACATTCGCTGCTTCTGAGTCTGCGGATGGGTGGAACACCTGGCTTCCCAACT 121748
QY 1213 GAATGGTTGATCGCATACCCCGACCGTTCGGCGGAAATGCCAAGAGCTCAACGGGG 1272
Db 121747 CCATGGTGGACCGCATCACCC---CTGAACACACCGACCGCGATGACATCAAGG 121691
QY 1273 CCAGTGGGCTGATCAGCAACCTGCGCGTGGTGGCGGAGATTTCCATCAGTGGGTGCTGG 1332
Db 121690 AATCGGCTTACATCATCGCTGGCGAGTGGTTCTGAAGATTTCAACCAATGGGTCTCG 121631
QY 1333 AAGACAGTTTCCGATGCGCGCTCGCGCTTGAAGAAACCGCGCTGAGATGGTGGGG 1392
Db 121630 AGGATGCTTTCACCGAGCGCGCGCTGACGAGAGGTGGCGTGCAGTCTGCTCG 121571
QY 1393 ACGTGACGAGTGGGAGTACGTCAGATCCGAATCTCAATGACGGGATGTCATGCTCT 1452
Db 121570 ACGTGAGCGCTTATGAATTAATGAAGTGGCGCTGCTCAACGCCCTCCACACGAGCTT 121511
QY 1453 GCTTCCAGGATTTCTGCTGCTGATGAGATGGATGAGCGCATGAGCGCATGGAACACGGAAC 1512
Db 121510 GCTACTTCCGCGACTTGGCTGGCCACCATGCTCCAGACGCTCATGCGGATACCGCT 121451
QY 1513 TCTTTGGCAATCTGAAGAATATCTCAACAGGATGTCATCCGACCGCTGAGCGGCTT 1572
Db 121450 TCCAGATTTCTCTGCTTACATGAGCGGAGACCGCCCTACCCTCAAGGAACCTC 121391
QY 1573 CAGGATGACGCTCGAAGGCTATCGGACAGCGCTCATCGCGTTCCTCCACAGGGA 1632
Db 121390 CAGGTGTCGATCTAGATGCTTATCGACGCCAACTCATCGCGATTCGCGACCGCGAG 121331
QY 1633 TGTCCGACAGCGCTCCGATGCTAGGATGCTGATGAGTGTTCAGGTTTCTGGA 1692
Db 121330 TCAAGACACCGCTACCGCGCTGTGTGGGAATCTCCGACCGCATTCCAAAGTGGCTG 121271
QY 1693 CGGAACCGTGTGCGGCGATCGAAGACAGCGGACCTGTACGTATAGGTTTCGGA 1752
Db 121270 TCCAGTCTAGCGGAAACCTCGCAGCGCGCGAGCGTCACTTTCTGACGCCATCG 121211
QY 1753 TTGCATCTATCTCGAAATGCTGCGTGGTTCGCGAGGAGGCG 1796
Db 121210 TCGCATCTGGCGCGCTACGAGAGGACCGCGAGCGAGGCG 121167

RESULT 3
US-09-802-208B-1
; Sequence 1, Application US/09802208B
; Publication No. US20030041352A1
; GENERAL INFORMATION:
; APPLICANT: Parrott, Wayne
; APPLICANT: LaFayette, Peter
; APPLICANT: Kane, Patrick
; TITLE OF INVENTION: Arabitol or Ribitol As Positive Selectable Markers
; FILE REFERENCE: USA-855R
; CURRENT APPLICATION NUMBER: US/09/802,208B
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1848
; TYPE: DNA

; ORGANISM: Escherichia coli
US-09-802-208B-1
Query Match 2.6%; Score 105; DB 9; Length 1848;
Best Local Similarity 46.8%; Pred No. 4.5e-22;
Matches 509; Conservative 0; Mismatches 560; Indels 18; Gaps 5;
QY 622 TGCATTCGGTGTAGTAACTTTTCGAGCCCATCAGGCGTCTACGTCGAGCATTC 681
Db 503 TGCACATCGGTTAGTCTTCTTTTCATCGCACATCAGCGTGGTATCTACACCGTTTC 562
QY 682 TTGAACACGCTCCGCACTGGCGATTTGTTGTTGGCTGACGGCGAGTGCACCTCAA 741
Db 563 AGGTGATGGCGGATAAAGCACTGTGAGCATTTGTCGGGCAATATTCTGAATGA---TGCTG 619
QY 742 AGAAAAAGCCGAGCAATTTCAAGGCCCGAGACTGCTGATTATTCCTCGACCGAGCGGCTC 801
Db 620 RACATGTCGTACAGCACTCAGTGCACAGAAAGTGCCTATGCTGTAACAAACCGTCAGCC 679
QY 802 COTCGGCAAGACGACGCTGCGCTGATGGCGCTGCTGAGTATCTGTTGCCCGG 861
Db 680 CGGAAGGGTAAAGCAATATGAAGAGATCACCTCAATTGAGAAAGTTGATACCGTGGCAGG 739
QY 862 CCGATCCGGAAGCCGCTGCTGAGCATCTTGTGATCCGCGCATCCGCTATCTGTTCCATGA 921
Db 740 CAGATTTACAACCGCTGATTTGCTGAAGGGGAGATCCGAAGACAAAGTGTGTTCTTCA 799
QY 922 CGATCAGCAAGCGGCTTACACATCAACGAGACGACCGGTGCTGCTGATCTGGAGATG 981
Db 800 CCGTCAACGAGCGGCTACTACCTGA---ATACCACTCAAACTGGAAGTTACAAATC 856
QY 982 CGGAGTAAAGCGCGACCTCAAGAACCCGGAAGCCGCTCTACCGTTTTCGTTACGTGG 1041
Db 857 CTGATTTAGCGCGACATCTTAAAGGGGATGCAAAACAATTTACGGTGTATTACCGGTA 916
QY 1042 TCGAGGCCCTCGGCTCGTGGGATGCGGTGTTAAGGCATTTACGGTATGCTGCTGTG 1101
Db 917 TCTTGAAGCGGATGCAAAATAACCGCG---ACCATAACCTGCTGAAATGCG 970
QY 1102 ATAACTCGCTCATACCGCAATTCGCGCGCAAGCGCTTCTCGCTGATGCGAAG---G 1158
Db 971 ATAAGTGGCGCAATTAATGCTGACCTTTCATGATGGCTGCTGTTGAGTTTCTCCAGCTAA 1030
QY 1159 CGCGGATCCGAGTGTGCGAAGTGTGAGGAAACCGGACCTTCCGCAACGGAATGG 1218
Db 1031 CTGCAAAACAGGATCTCATCGACTGCTGAGTACAAATACCACTTGCCTGCAATACCATGG 1090
QY 1219 TTGATCGCATCACCCCGACCGTTTCGCGGAAATCGCAAGAGCTCAACCGCGCAGTG 1278
Db 1091 TTGACCGCATTACGCTCTCGCGCAGCAGAACTTCCGCGCAGCATCAAGGCTCAAGCGG 1150
QY 1279 GGCTGATCAGCACTCGGCTGCGCGGAGGATTTCCATCAGTGGGTGCTGGAAGACC 1338
Db 1151 GTATTGCGGATTAAGCGCGCGGTAATGCGCGAAACCTTTATCCAGTGGGTGCTGGAAGATA 1210
QY 1339 AGTTTGGGATGCGGCTCCGCGCTTGAAAGCCGCGCTGAGATGGTGGGAGCTGA 1398
Db 1211 ATTTCCGCTGATGCTCGTCCGCACTGGAAGGTCGGTGTGCGAAGTGGTGGCTCGGTA 1270
QY 1399 CGGACTGGGAGTCAAGATTCGAATGCTCAATGACGGGATGCTCATGCTCTGCTTCC 1458
Db 1271 TCCCTATGAAGGCGAAGATTCGCAATTTCACTTCACTTCAACAGTGTGCTGCTGGG 1330
QY 1459 CAGGATTTCTGCTGCTGCTGATGAGATGGATGAGCGCATTCAGACAGCGCAACCTCTG 1518
Db 1331 CAGGTACCTTAACTCGTCAAAATATATCCAGAAAGCAATACGAGTTTATCTATC 1390
QY 1519 GCAATCTGAAGAACTATCTCAACAGGATGCTATCCCGACCTGAAGCGCTTCAAGCA 1578
Db 1391 AGATTGCGCGCGCTACCTGACGGAAGATGCTATCTCTTGT---GGCGGATACGTA 1447
QY 1579 TGACGCTCGAAGGCTATCGGACAGCGTCACTACGCGTTTCTCCAAAGCGGATGCTCG 1638

Db 1448 TCGATTTCCCAACTACCGTATGTTGTAATCAAGCGTTTACCAATCCACATATTCAGG 1507
Qy 1639 ACCAGAGCTCCGATGCTAGCATGCTTCCAAAGTTCAGGTCTTCTGGACGGAAA 1698
Db 1508 ACACCAACCAACGCGTGGTGGATGTTTCTCGAAAAATCCGGCGATGATTCGCCCA 1567
Qy 1699 CCGTGGC 1705
Db 1568 CACTGGC 1574

RESULT 4

US-09-789-561-27
; Sequence 27, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 2371
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-789-561-27

Query Match 1.0%; Score 42; DB 10; Length 2371;

Best Local Similarity 50.0%; Pred. No. 0.027;

Matches 105; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 1157 GCGCGCGATCCGGAGTGGCGAAGTGGATTGAGGAAAACCGGACCTTCCGGAACGGAAT 1216
Db 527 GCGAGCCGGCTGGGCTGTTGTCATGCTTCCAGAACCTGTGTCATGTTGATGAGGACTT 586
Qy 1217 GTTGTATCGCATCACCGACCGCTTTCGGCGGAATCCCAAGAGCTCAACGCGGCCAG 1276
Db 587 TGTGGACTGGTTCATCCCGGACATCCCAAGGACATCAGCCAGCAGATCCCAAGGAGAA 646
Qy 1277 TGGGTGATGATGACGACCTGCGCTGGTGGCGGAGGATTTCCATCAGTGGTCTGGAAGA 1336
Db 647 GGTGCTCATGTTGGAGCTGTTTCATCGCGGAGGAGCAAGCAAGCAGCAGCTGCTGGAAC 706
Qy 1337 CCAGTTTCCGATGCGCTGCGCGCTTGA 1366
Db 707 CTGGATGGAGAGGAGCGGACGAGGACGA 736

RESULT 5

US-09-822-846-491/C
; Sequence 491, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.

; APPLICANT: Clark, Hillary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 491
; LENGTH: 2614
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-846-491

Query Match 1.0%; Score 41.6; DB 9; Length 2614;

Best Local Similarity 51.3%; Pred. No. 0.037;

Matches 121; Conservative 0; Mismatches 114; Indels 1; Gaps 1;

Qy 3337 CTGATGGATCTTTTGATCAAGCGTCTCCCAATGTCGCAAGAGTGTGAACTGGTTG 3396
Db 2172 CTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2113
Qy 3397 CGGGAACGGATCAATAATGGTTCCGCGCGGCGTCCGCTGCTGCTGCTGCTGCTGCTG 3456
Db 2112 CCGCAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2053
Qy 3457 GTTCTGTCCATCCGCTGTTCTGGTCTGTGGATGCTGCGGTCGCGGTGATGTTGCTT 3516
Db 2052 TTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1993
Qy 3517 GCGCAGGATATTCGTTCTTCCGTCGCTTCCAGGCGCGCTTTCGCGCTGGATCGA 3572
Db 1992 AGGTAGGATGG-CTGTTGAGGAGCGCTGCTGCGACGCTTGGGCACTCATTTGA 1938

RESULT 6

US-09-822-849A-393
; Sequence 393, Application US/09822849A
; Patent No. US20020045170A1
; GENERAL INFORMATION:
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hillary
; APPLICANT: Fechtel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakara
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6403
; CURRENT APPLICATION NUMBER: US/09/822,849A
; CURRENT FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: 60/195,582
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 598
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 393
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-849A-393

Query Match 1.0%; Score 41.4; DB 10; Length 1512;

Best Local Similarity 48.5%; Pred. No. 0.034;

Matches 114; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

Db 560 ACGCGTGGCGCTTTAAACAAGGGTTATCGATTGGCTTGAAGAGCAAAAGCTGTTCGA 619
Qy 1210 ACGGAATGGTTGATCGATCAACCCGACCGCTTCGCGGGAATCGCCAAGAGCTCAACG 1269
Db 620 ATACGCTTGTGGACAGGATCGTACCGGATTCGCCGAAGAAGAAATGCACACCGCTTGAAG 679
Qy 1270 CGGCGAGTGGCTGGATGACGACCTGCGCGTGGTGGCGGAGGATTTCCATCAGTGGGTGC 1329
Db 680 AGCGCGCGGCTATCAGCATTCCTTGATGGTGAAGGCGGACCGCTATCATTTGTCGTGA 739
Qy 1330 TGAAGACCACTTTGGCGGATGG----CGTCCGCGCTTTGAAAAAGCCGG-----CGTGC 1380
Db 740 TTGAAGGCGCAAAATGGCTTCGAGACGAGCTTCGCGCTTCATGAGCGCGCTGTAATGTG 799
Qy 1381 AGATGGTGGGAGCTGACGAGCTGGGAGTACGTCAAGATCCGATCCGATGCTCAATCGAGGC 1440
Db 800 AGTTGTCCAGGATGTGACGCGGTACAGGATGAAGAAAGTGAAGATTTAAACGCGCGCC 859
Qy 1441 ATGTCATGCTCTGCTTCCAGGCAATCTGTCGCTATGAGAATGTGATGACGCCATTTG 1500
Db 860 ATACCGGATGACCGCGTCTCTTATTTGGCAGGCTTGAGACGTTTCGCGAAGCGGTG 919
Qy 1501 AAGACGGAAGTCTTGGCAATCTGAAGAATCTATCTCAACAAGAGTGTATCCCGACCC 1560
Db 920 AAGA---CGAGCTTATCGGCTCATTTGTGAAAGAGCTGATCCAGCGGAGGTGCTCGAGA 976
Qy 1561 TGAAGGCGCTTACGAGTATGAGCTCGAAGCTATCGGACGAGCTCATCAGCGCTTCT 1620
Db 977 CGATCGATATCCCGATAGCAGGCTTGACAACTAGTGGGAAGAGATTCGACCGCTTA 1036
Qy 1621 CCAACAAGCGATGTGCGACGACGCTCGGATTTGCTAGCGATGGCTGTTTCCAAAGTTC 1680
Db 1037 AAAACCGCTTTATCCGTCACAAGCTGATCGATATTTCTTAAATCTCTTTCCAAAATCA 1096
Qy 1681 AGGT 1684
Db 1097 AAGT 1100

RESULT 10
US-10-152-040-26/c
; Sequence 26, Application US/10152040
; Publication No. US20030077251A1
; GENERAL INFORMATION:
; APPLICANT: ESCRIOU, NICOLAS
; APPLICANT: VAN DER WERF, SYLVIE
; APPLICANT: VIGNUZZI, MARCO
; APPLICANT: GERBAUD, SYLVIE
; TITLE OF INVENTION: REPLICONS DERIVED FROM POSITIVE STRAND RNA VIRUS
; TITLE OF INVENTION: GENOMES USEFUL FOR THE PRODUCTION OF HETEROLOGOUS
; FILE REFERENCE: 03495.0229-00000
; CURRENT APPLICATION NUMBER: US/10/152.040
; PRIOR FILING DATE: 2002-06-27
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 8017
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-152-040-26

Query Match 1.0%; Score 40.4; DB 9; Length 8017;
Best Local Similarity 97.6%; Pred. No. 0.14;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAGCTTGCATCCCTGCAGGTCGACTCTAGAGGATCCCGTTTTT 42
|||||

Db 4872 AAGCTTGCATCCCTGCAGGTCGACTCTAGAGGATCCCGTTTTT 4831
RESULT 11
US-10-152-040-27/c
; Sequence 27, Application US/10152040
; Publication No. US20030077251A1
; GENERAL INFORMATION:
; APPLICANT: ESCRIOU, NICOLAS
; APPLICANT: VAN DER WERF, SYLVIE
; APPLICANT: VIGNUZZI, MARCO
; APPLICANT: GERBAUD, SYLVIE
; TITLE OF INVENTION: REPLICONS DERIVED FROM POSITIVE STRAND RNA VIRUS
; TITLE OF INVENTION: GENOMES USEFUL FOR THE PRODUCTION OF HETEROLOGOUS
; FILE REFERENCE: 03495.0229-00000
; CURRENT APPLICATION NUMBER: US/10/152.040
; PRIOR FILING DATE: 2002-06-27
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 8092
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-152-040-27

Query Match 1.0%; Score 40.4; DB 9; Length 8092;
Best Local Similarity 97.6%; Pred. No. 0.14;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAGCTTGCATCCCTGCAGGTCGACTCTAGAGGATCCCGTTTTT 42
|||||
Db 4947 AAGCTTGCATCCCTGCAGGTCGACTCTAGAGGATCCCGTTTTT 4906

RESULT 12
US-10-152-040-28/c
; Sequence 28, Application US/10152040
; Publication No. US20030077251A1
; GENERAL INFORMATION:
; APPLICANT: ESCRIOU, NICOLAS
; APPLICANT: VAN DER WERF, SYLVIE
; APPLICANT: VIGNUZZI, MARCO
; APPLICANT: GERBAUD, SYLVIE
; TITLE OF INVENTION: REPLICONS DERIVED FROM POSITIVE STRAND RNA VIRUS
; TITLE OF INVENTION: GENOMES USEFUL FOR THE PRODUCTION OF HETEROLOGOUS
; FILE REFERENCE: 03495.0229-00000
; CURRENT APPLICATION NUMBER: US/10/152.040
; PRIOR FILING DATE: 2002-06-27
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 10417
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-152-040-28

Query Match 1.0%; Score 40.4; DB 9; Length 10417;
Best Local Similarity 97.6%; Pred. No. 0.16;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 AAGCTTGCATCCCTGCAGGTCGACTCTAGAGGATCCCGTTTTT 42
|||||

OTHER INFORMATION: ORF 24; negative strandedness
NAME/KEY: misc.feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
NAME/KEY: misc.feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
NAME/KEY: misc.feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
NAME/KEY: misc.feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
NAME/KEY: misc.feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc.feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc.feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
NAME/KEY: misc.feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
NAME/KEY: misc.feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only

US-09-976-059-1

Query Match
Best Local Similarity 1.0%; Score 40.2; DB 9; Length 88421;
Matches 96; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 3445 ATGCTGGCGGGTCTCTCCATCTCGCTGTCTGCGGTCTGTGGAGCTGCCGGTCGGC 3504
Db 37184 ACGGTCCGGCGAGCGGATCTCCGGCGCATCCCGCGAGGTGTCCGCCAGTAGGC 37125
QY 3505 GTGATGTTCTGCGCAGATATCCGTTCTTCGTCCGCTTCAGGCGCGCTCTTCGCG 3564
Db 37124 GAGCTGGCGCTCGCCAGCGTCTCCGGGTACCGCGCGCGCGCCAGCAACTCTCTCTGCCA 37065
QY 3565 TGGATCGAAGCTCAACATCCGATGCTGGGCGCTTCGCGGGAAGCGGAGAGCTAA 3624
Db 37064 GAGCGCGTAGTCGGGTACTCAGCGCGCAGCGGCTCCCGCGCGGTGCCCGCGCCAG 37005
QY 3625 CGTTCGTC 3633
Db 37004 CCGGCCTC 36996

RESULT 15
US-09-981-353-85/c
; Sequence 85, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 85
; LENGTH: 3115
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 1102315.3
; NAME/KEY: unsure
; LOCATION: 2713, 2719, 3094

OTHER INFORMATION: a, t, c, g, or other
US-09-981-353-85

Query Match
Best Local Similarity 1.0%; -Score 40; DB 9; Length 3115;
Matches 130; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 2628 GAACAGTCCGTCATGAGTTTCAAGCGGCGCGTATTATCGCATAGGCTTGGCCCAT 2687
Db 773 GCACAGTGGTAGATGAAGGACCCAGGCGCTCTTGTCACTGACATGCTCCGTGACCCAC 714
QY 2688 TTCGGGGCATACATCTCGAAATCGTCCGTCGCGGCGCGGATCGAACAGCATGCC 2747
Db 713 CATCTCCTCTGGATGATGTAGCTGCGGCAGGTAGTCCCGCTCTTTCACGTTTCATGAG 654
QY 2748 GACTTCCTTGTGTATTCGGGGGGAAGTGGAAAGAGGTCTTGAAGCGTTGATTTCTGTG 2807
Db 653 GAGCTCCAGAAAGTTGCGAGGGGCGAGCAATGTTGTTGAGTTTCGATGACATAGCA 594
QY 2808 TCGGTTCAACGCGCGCTTCGCGCAGCTTCGCGCAGGCGCAACAGGCGCATGCGC 2867
Db 593 CTTGTCCAGGAGATATCATGTAGCAGTCCAGCCCGCTGGAGTCTATGATGATGTC 534
QY 2868 GTAAAGCTGATCTCGTTTTCGCGGCGCGGAGCAATCTTG 2907
Db 533 TGCAGGCTCACCGCGCGCAAACTGGGGCACAGGCGCGTTG 494

Search completed: May 18, 2003, 21:32:28
Job time : 3984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 21:32:35 ; Search time 5382 Seconds
(without alignments)
12382.837 Million cell updates/sec

Title: US-09-926-163B-1

Perfect score: 4115

Sequence: 1 aagcttcacgtcctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pin:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	180.8	4.4	801	17 BH395246	BH395246 AG-ND-144
C 2	155.4	3.8	761	17 AZ139030	AZ139030 SP.0178_B
C 3	146.8	3.6	840	10 BE641454	BE641454 Cri2_3_F0
C 4	131.8	3.2	799	17 CNS010W5	AL156645 Anopheles
5	88.6	2.2	527	9 A965891	A965891 o8g06al.r
6	86.6	2.1	481	17 AZ049105	AZ049105 GSSBru050

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

7	85.4	2.1	624	10	AM671891	AM671891 LG1_352_G
8	80.4	2.0	519	13	BI191648	BI191648 11603fs.r
9	80.4	2.0	519	13	BI191768	BI191768 12603fs.r
10	77.6	1.9	520	17	BH143537	BH143537 TDGEB13TH
11	69.4	1.7	556	17	AQ500091	AQ500091 V34C6 mTn
12	54.2	1.3	506	12	BG608411	BG608411 296895 MA
13	51.4	1.2	788	17	AZ933800	AZ933800 BJ__Ba000
14	49.2	1.2	1101	17	CNS017SY	AL108460 Drosophil
15	48	1.2	861	9	AL539916	AL539916 AL539916
16	47.6	1.2	927	12	BE733079	BE733079 601569589
17	47.2	1.1	244	17	AZ578838	AZ578838 28307 Sho
18	47	1.1	830	14	BQ672858	BQ672858 AGENCOURT
19	47	1.1	934	14	BQ641340	BQ641340 AGENCOURT
20	47	1.1	1012	13	BM549784	BM549784 AGENCOURT
21	46.4	1.1	812	13	BI600315	BI600315 603250632
22	46.4	1.1	831	12	BG818263	BG818263 602779895
23	46.4	1.1	861	14	BQ437338	BQ437338 AGENCOURT
24	46.4	1.1	875	13	BM009512	BM009512 603630050
25	46.4	1.1	882	13	BI856423	BI856423 603387791
26	46.4	1.1	898	12	BG470953	BG470953 602511926
27	46.4	1.1	922	12	BG754237	BG754237 602709788
28	46.4	1.1	925	17	CNS0091P	AL053013 Drosophil
29	46.4	1.1	992	14	BQ057936	BQ057936 AGENCOURT
30	46.4	1.1	1016	13	BM562009	BM562009 AGENCOURT
31	46.2	1.1	824	12	BG831754	BG831754 AGENCOURT
32	46.2	1.1	1030	13	BM563059	BM563059 AGENCOURT
33	46	1.1	121	17	AZ579022	AZ579022 30b10 Sho
34	46	1.1	912	14	BQ950903	BQ950903 AGENCOURT
35	45.8	1.1	897	14	BQ933488	BQ933488 AGENCOURT
36	45.6	1.1	630	13	BJ272413	BJ272413 Drosophil
37	45.6	1.1	925	17	CNS0091P	AA383457 EST96834
38	45.4	1.1	221	9	AA383457	BG756596 602713747
39	45.4	1.1	770	12	BG756596	BG756729 602715595
40	45.4	1.1	788	12	BG756729	BM048024 603620340
41	45.4	1.1	885	13	BM048024	BB643958 BB643958
42	45.2	1.1	681	10	BI929012	BI929012 au64all.y
43	45.2	1.1	786	9	A1929012	BQ231485 AGENCOURT
44	45.2	1.1	864	14	BQ231485	BG764520 602736565
45	44.8	1.1	765	12	BG764520	

ALIGNMENTS

RESULT 1
BH395246/C
LOCUS BH395246 801 bp DNA linear GSS 11-DEC-2001
DEFINITION AG-ND-144F15.TR ND-TAM Anopheles gambiae genomic clone AG-ND-144F15
ACCESSION BH395246
VERSION BH395246.1 GI:17341387
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 801)
AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B. J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSS: AG-ND-144F15.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae

to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.
Seq primer: M13 Rev
Class: BAC ends.

FEATURES

Source
1. 801
Location/Qualifiers

/organism="Anopheles gambiae"
/strain="P8T"
/db_xref="taxon:7165"
/clone="AG-ND-144F15"
/clone.lib="ND-TAM"
/note="Vector: pECBAC1; Site_1: HindIII"
210 a 162 c 156 g 273 t

BASE COUNT

210 a 162 c 156 g 273 t

Query Match 4.4%; Score 180.8; DB 17; Length 801;

Best Local Similarity 55.4%; Pred. No. 2e-36;

Matches 350; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

QY 867 CGGAAGCCGCTGTAAGCATCTTGTGATCGGCCATCCGATGTTTCATGACGATC 926

Db 11

Db 634 CCACAGGATCTATTGAATAAATAGCTGACAGCAACACCAATATTACACTAACATC 575

QY 927 ACAGGAGCGGCTACAAATCAACAGACGACCGGTGCTGATCTGGAAATCGGCA 986

Db 11

Db 574 ACTGAAGTGGCTATAATCTGGATAAGCGCAACAGAAATTTATCTGAATAAGAAAT 515

QY 987 GTAAAGCGGACCTCAAGAACCGGAAAGCGCTTACCGTTTTCGGTTAGTGTGAG 1046

Db 11

Db 514 ATACAATGACCTGAAGAACCTGAATTCGCACTACTGTTTCGGTTTTCATTCGCGAA 455

QY 1047 GCCCTGCTGCTGTTGGGATCGCCGGTGGTAGGCAATTTAGGTCATGCTCTGTGATAAC 1106

Db 11

Db 454 GGTCTCTGCTCGGAAATCCAAAGCAATGGCAGTATTACCAATTTCTCTCTGTGACAAAC 395

QY 1107 CTGCGTCATACGCGCAATGTGCGCGCAAGCGCTTCTCTCGGTATGCGAAGCGCGCAT 1166

Db 11

Db 394 CTTCAACATACGGCAATGCTGCGCAAAATGCAATTTCTCATTTATTGAAGCTCAGAT 335

QY 1167 CGGAGTTGGGAGTGGATTGAGGAAACGCGACCTTCCGGAAGCAATGTTGATCGC 1226

Db 11

Db 334 AAAGATTGGGAGATGGTTGAGCAAAATGTAACCTTCCCAACAGTAGTAGACAGA 275

QY 1227 ATCAACCCGACCGTTTCGGGGAATCCCAAGAGCTCAACGGCGGCGAGTGGCTGGAT 1286

Db 11

Db 274 ATAACACCTGTAACTACAAATAGAGATGTAAGAGGCTTAAACGAAAGTGGTATCGCA 215

QY 1287 GACGACCTGCGCTGCTGCGCGAGGATTTCCATCAGTGGTGTGCTGGAAGACAGTTTCGG 1346

Db 11

Db 214 GATAGGCTCGGTATATTGGAAGATTTTACCCTAATGGTAAAGAACAAATTTTATT 155

QY 1347 GATGCCCTGCGCCCTTTGAAAAGCGCGCTGACAGTGGTGGGAGCTGACGACCTGG 1406

Db 11

Db 154 GCGGAAGACCTCGCTGGAAGAGTGGGTTACTTTTCAAAAAGATGTAACGCTTAC 95

QY 1407 GAGTACGTCAGATCCGAATGCTCAATGACGGGCAATGTCGCTGTTCCAGGCAAT 1466

Db 11

Db 94 GAAAAATGAAGCTGAGTCTGCTAAATGCTTCTCATACCTTCTATCCTTCAATTT 35

QY 1467 CTGTCGCGCTATGAGAATGTGATGACGCCAT 1498

Db 11

Db 34 TTAGCCGATACCGCAAAAGTAGCAAGCCAT 3

RESULT 2

AZ139030/c

LOCUS

DEFINITION SP_0178_B2_F03_r7A Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=178 Col=6 Row=L, DNA sequence.

ACCESSION

AZ139030.1 GI:8290933

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Strongylocentrotus purpuratus.
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidea; Echinoidea; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
1 (bases 1 to 761)
Cameron, R.A., Mahairas, G., Rast, J.P., Martinez, P., Biondi, T.R.,
Swartzell, S., Wallace, J.C., Poustka, A.J., Livingston, B.T., Wray,
G.A., Ettensohn, C.A., Lehrach, H., Britten, R.J., Davidson, E.H. and
Hood, L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
20402566
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 178 row: L column: 6
Seq primer: T7
Class: BAC ends
High quality sequence stop: 761.
Location/Qualifiers
1. 761
/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=178 Col=6 Row=L"
/clone.lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="organ: sperm; Vector: BACE3.6; BAC Clones in E-Coli
DH10B"

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Conservative

Mismatches

Indels

Gaps

Score

Pred. No. 1e-29;

Length

761;

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

QY

Db

```
Db 315 GTTCGGACGCTCGGAATGGAAGGAGGAGCCGAACCTGGTAGCGATGCTGCCT 256
Qy 1404 TGGAGTACGTCAAGATCCGAATGCTCAATGACGGGATGTCATGCTCTGCTTCCAGGC 1463
Db 255 TATGAAGAGATGAAGTGGCGATGCTCAACGGCAGTCAATTCATTCCTGGCGTATCTGGG 196
Qy 1464 ATTCTGGTGGCTATGAGAATGTGGATGACGCCAATGAAGA 1504
Db 195 TATCTTGAGGATATACGACATTAATGACTGTATGAAGA 155

RESULT 3
BE641454 840 bp mRNA linear EST 01-SEP-2000
LOCUS Ceratopteris Spore Library Ceratopteris richardii
DEFINITION cDNA clone Cr12_3_F06 5', mRNA sequence.
ACCESSION BE641454
VERSION BE641454.1 GI:9959119
KEYWORDS EST.
SOURCE Ceratopteris richardii.
ORGANISM Ceratopteris richardii.
REFERENCE 1 (bases 1 to 840)
AUTHORS Chatterjee,A., San Miguel,P., Stout,S.C., Banks,J. and Roux,S.J.
TITLE Expressed sequence tags of cDNA clones from a C. richardii library
JOURNAL Unpublished (2000)
COMMENT Contact: Roux SJ
Section of Molecular Cell and Developmental Biology
University of Texas
Biology Building, Room 16, Austin, TX 78712, USA
Tel: 512 471 4238
Fax: 512 232 3402
Email: sroux@uts.cc.utexas.edu
Plate: Cr12.3 row: F column: 06
Seq primer: Sp6.
Location/Qualifiers
1. 840
/organism="Ceratopteris richardii"
/cultivar="Brogn"
/db_xref="taxon:49495"
/clone="Cr12_3_F06"
/clone_lib="Ceratopteris Spore Library"
/tissue_type="Gametophyte"
/dev_stage="20 hours after germination initiation"
/note="vector: pCMVSPORT6; EST sequence from cDNA library.
cDNA library constructed from mRNA isolated from C.
richardii spores that had developed for 20 hours after
their germination had been initiated by white light."
BASE COUNT 168 a 313 c 226 g 133 t
ORIGIN

Query Match 3.68; Score 146.8; DB 10; Length 840;
Best Local Similarity 50.68; Pred. No. 2e-27;
Matches 406; Conservative 0; Mismatches 392; Indels 4; Gaps 2;

Qy 576 GTCCAGGCTCCCTATGACATGACATGCGGGATCAAGCCTGGGATCGTCGATTCGGGTGA 635
Db 43 GTCAAAATCCGACCTAGCGGAGGATGACGTACGAGGAGGAATGTCCAGCTCGGGCTC 102
Qy 636 GGTAACATTTTTCGAGCCCATGAGCGCTTCTACGTGAGCAGATTTTGAACAGCTCCG 695
Db 103 GGCGGCTTCCACAGAGCCCATCTGGCGCTCTACGTGACAAATTTGATGCAACACCGGC 162
Qy 696 GACTGGCGGATTTGGTTGGCTTACGGGCGAGTACCGTTCAAGAAAGAAAGCCGAG 755
Db 163 CACCGCGACTGGGCGCATCTGGCGGCTGGCGTGGACCCCAACGACGCGGCATCGCGAC 222
Qy 756 GAATTAAGGCCCGAGGACTGCTGATTCCTGACGAGAGCGGCTCCGTCGGCAAGAGC 815
Db 223 GTGCTCAGGCCCGAGGACCACTGTACACCGTATCAGCGGCTCCGCAAGGCG---AGC 279
```

```
Qy 816 ACGTGGCGGTGATGGGCGGCTGCGTGAATATCTGCTTGGCCGGCGGATCCGGAAGCC 875
Db 280 TTCGCCAACGCTGTCGCGAGCATCAACTCGTTCTTTCGCCCGCGAGCAGCCGAGGCC 339
Qy 876 GTGCTGAAGCATCTTGTGATCGGCCATCGGCATGTTTCCATGACGATCACCAGAGGC 935
Db 340 GTCATTGCCAAGATGGCGCACCCCGACACCATCTCTCCCTGACCATCACCAGAGC 399
Qy 936 GGCTACAACTCAACGAGACGCGGTTCGATCTGGAGAATGGGGAGTAAAGGCC 995
Db 400 GGCTACTACTACACGAAACACCCAGAGCTGAGGCCAACCATTCGGACATCCAGCAC 459
Qy 996 GACCTCAAGAACCCGAAAAAGCCGCTACCGTTTTTCGTTTACGTTGGTGGAGCCCTCGGT 1055
Db 460 GACCTGGCCAAAGAGAACGCCCGCTCAGCACCTTTGGCTTCTTACGCCGCGCTGGCC 519
Qy 1056 CGTCTTGGGATGCGGTTGAAGCATTTACGGTCATGCTCTGTATACCTCGGTCAT 1115
Db 520 AAGCGCCACGCGAAGGCGCTTCAAGCCCTTCAACCGCTATGCTCTCGCACACATCGAAG 579
Qy 1116 AACGGCAATGTCGCCCGCAAGGCGCTTCTCGGCTATGCGAAGGCGCGGATCCGGAGTTG 1175
Db 580 ACGGGTCCATCACCCCGCACATGCTCGAGTCGTTTGGCCCGGTCGCCAACCCGAGCTG 639
Qy 1176 GCGAAGTGGATTGAGGAAAAACCGGACCTTCCGAAACGAATGTTGATCGCATCACCCCG 1235
Db 640 GCC-AGTGGATGCGCCAGGAGGTGCTTCCCAACACGCTTGGGCGGACCGCATACACC 698
Qy 1236 ACCGTTTGGGGAAATCGCCAAGAAGCTCAACCGCGCAGTGGGCTGGATGAGCAGCTG 1295
Db 699 CGCACCTTCCAAAACGACCTTTGAGTCTTGGCCAGAGCTTTGGCCCTCGAGAGCGCTGG 758
Qy 1296 CCGCTGTGCGCGAGGATTTTCATCATGCTGGGTGCTGGAAGACCACTTTGCGGATGGCGGT 1355
Db 759 CCGCTTGTCCACCAACCTTCATGTCATGGGGTCTTTGAGGACACASTTCAGCAAGGCC 818
Qy 1356 CCGCGCGCTTGAAGAACCGCGG 1377
Db 819 CCGCGGTTTCAAGGTTGGGG 840

RESULT 4
CNS01QW5/c
LOCUS Anopheles gambiae GSS SP6 end of clone 3LM20 of NotreDamel library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL156645
VERSION AL156645.1 GI:7017564
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoides;
Anopheles.
REFERENCE 1 (bases 1 to 799)
AUTHORS Direct Submission
TITLE Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
REFERENCE 2 (bases 1 to 799)
AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BIMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur.
FEATURES Location/Qualifiers
1. 799
source
```

/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="31M20"
/clone_lib="NotreDamel"
/note="end : SP6"

BASE COUNT 151 a 245 c 206 g 167 t 30 others
ORIGIN

Query Match 3.2%; Score 131.8; DB 17; Length 799;
Best Local Similarity 51.6%; Pred. No. 1.8e-23;
Matches 332; Conservative 6; Mismatches 302; Indels 4; Gaps 3;

```
QY 861 GCCGATCCGGAAGCCGCTGTAAGCATCTTGTGTGATCGCGCATCCGCATCGTTTCCATG 920
DB 649 GACGGCTCGAGACCGCTCTGCCCGGATGCGGACCGCAGGTGGCGATGTGTCGTG 590

QY 921 ACGATCCAGGAGCGGCTACAAATCAACAGACGACCGGTGCTTCGATCTGGAGAAT 980
DB 589 ACCATCCAGGAGAAAGGCTACTGCCATTTCCCGCGGACCGGCAACTGATCTCGATCAC 530

QY 981 CGGCAGTAAGCGCGACCTCAAGNACCCGGAAGCGCTTACCGTTTTCGGTTACGTG 1040
DB 529 CCTTTTATCGTCGACAGCTGMAAAACCCGACCAAGCGAAATCTGCCCGGCTGTGTC 470

QY 1041 GTCGAGGCCCTCGCTGCTGTTGGATCGCGTGTGAAGCATTTTACCGTTCATGTCCTGT 1100
DB 469 GTGGAAGSCTCGCGCGGTAAGCGCGGCGCTGCC--GGCGTTTAGCGTATGTCCTGC 411

QY 1101 GATACTTCGCTATACGGAATGTCCCGCGAAGGCTTCTCGGTATGCGAAGCG 1160
DB 410 GACAAATCGCGGAGAACGGTCACGTGATGCGMAATGTACCTGCGCGTACGCGCGTGC 351

QY 1161 CGCATCCGAGTTCGCGAAGTGTGAGAAACCGACCTTCCCGAAAGGATGTT 1220
DB 350 GTAGACGTGAGCTGGCGAGTGGATCGAATCCACGTTTACCTTCCCGTCAACGATGTG 291

QY 1221 GATCCATCACCCGACCGTTCGCGGAAATCGCAAGAGCTCAACCGCGCGAGTGG 1280
DB 290 GATCGCATNGTGT--CGTCAACCCCGACGCTGGATAAATGAAGCTTACCGGC 232

QY 1281 CTGGATGACGACCTCGCGCTGGTGGCCGAGGATTTCCATCATGAGTGGTCTGGAAGACCAG 1340
DB 231 GTCCGCGATCGCGCGCGCTGTCGCAAGGCTGCAAGTCCCAAGGCTGCAAGATAAC 172

QY 1341 TTTCGGATGCGCGTCCCGCTTCAAAACGCGCTGCAGATGCTCGGCGAGTGCAG 1400
DB 171 TTGTGGCGCGCTCCAGCTGGGACAAAGCCGAGCGGAGCTGGTGTACAGCTGATC 112

QY 1401 GACTGGGAGTACGTCAAGATCCGAATGCTCAATGACGGGCTGTCATGCTCTGCTTCCCA 1460
DB 111 CGGTTTGAAGATGA--ATGCGCATGCTTAACGCGACCCACTMATTCTTGGCTACCTG 54

QY 1461 GCATTCGTGCGGTATGAGAATGTGATGACGCGCATTTGAAGA 1504
DB 53 GGCTATCTGGGGGCTATCAGMATATTAACGAATGTATGAAGA 10
```

AA965891 527 bp mRNA linear EST 31-JUL-1998
LOCUS o8g06al.r1 Aspergillus nidulans 24hr asexual developmental and
vegetative cDNA lambda zap library Emericella nidulans cDNA clone
o8g06al 5', mRNA sequence.

ACCESSION AA965891
VERSION AA965891.1 GI:3139775

KEYWORDS EST.

SOURCE Emericella nidulans.

ORGANISM Emericella nidulans

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Eurotiales; Trichocomaceae; Emericella.

REFERENCE 1 (bases 1 to 527)

AUTHORS Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,

TITLE
JOURNAL
COMMENT

Prade,R. and Roe,B.
An Aspergillus nidulans EST Database
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Seq primer: T3
High quality sequence stop: 405.

FEATURES
source

1..527
Location/Qualifiers

/organism="Emericella nidulans"

/strain="FGSC A26"

/db_xref="taxon:162425"

/clone="o8g06al"

/clone_lib="Aspergillus nidulans 24hr asexual

developmental and vegetative cDNA lambda zap library"

/tissue_type="vegetative mycelia, asexual structures"

/note="vector: pBluescript SK-; Site.1: EcoRI; Site.2:

XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript

3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 112 a 187 c 128 g 100 t

ORIGIN

Query Match 2.2%; Score 88.6; DB 9; Length 527;

Best Local Similarity 52.3%; Pred. No. 3.9e-12;

Matches 220; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

```
QY 593 CTCCTCCCTATGACATCGACGGGATCAAGCCCTGGATCGTGCATTCGGTGTAGTACT 642
DB 39 CTACCTACCCCTCGCAATGGCGCGCTCAAGGAGGTATTGTCCATATCGCGCGCTG 98

QY 643 TTTTTCGAGCCATGAGGCGTTCTACGTCGAGCAGATTTCTTGAACACACGCTCCGACTGG 702
DB 99 TCCACCGAGCCATTTGGCGCTCTATATCGACCGGCTGATCGAGNACCACCGGCTCACTG 158

QY 703 CGATTGTTGTTGGCTGACGGGCTGACGGGCTGACCGTTCAAGAAAAAGCCGAGGAATCA 762
DB 159 ACTATGCGATCGCGGCTGTTGCTGTGACCCCTTTGACAAGAAGATGCGCATATCTCA 218

QY 763 AGGCCCCAGACTCGCTGTTTCCCTGACCGGAGACGGCTCCGTCGGGCAAGACGCGTGC 822
DB 219 GGCTTCAAGACCACTCTACACCGTCTACGAGCGATCGGCCAAGGCG---AGCTTCGCCA 275

QY 823 CGGTCATGGGCGGCTGCGTGACTATCTGCTTGGCCCCGCGGATCCGGAAGCCGCTGCTGA 882
DB 276 ACGTGGTGGCTCGATCAACTCGTTCTCTTTGCTCTGACGACCGCGGAGCCGCTCGTCG 335

QY 883 AGCATCTTGTGATCCGGCCATTCGCGATCGTTTCCATGACGATCAGGAAGCGGCTACA 942
DB 336 CCAAGATGGCCACCGCGGACACACATCGTCTCCCTGACTATCACCAATATCCGGCTACT 395

QY 943 ACATCAAGGACGACGCGTGGTTCGATCTGGAGATGCGGCAAGTAAAGCCGACCTCA 1002
DB 396 ACTACATGAGAACACGACGACGAGCTCGTAAGCGAGCACCCCGGACATCCAGCTTTCAC 455

QY 1003 A 1003
DB 456 A 456
```

RESULT 6

AA965891

LOCUS

DEFINITION

ABORTUS genomic clone B4G18, DNA sequence.

ACCESSION

AA965891.1 GI:7273020

AZ049105 481 bp DNA linear GSS 01-SEP-2000

GSSBRU0507 Sheared genomic library Brucella melitensis biovar

ABORTUS genomic clone B4G18, DNA sequence.

AZ049105

AZ049105.1 GI:7273020

```

KEYWORDS
SOURCE
ORGANISM

REFERENCE
1 (bases 1 to 481)
AUTHORS
Sanchez,D.O., Zandomeni,R., Cravero,S., Rosetti,O., Grau,O. and
Ugalde,R.
TITLE
Gene discovery through genomic sequencing survey of the Brucella
abortus genome
JOURNAL
COMMENT
Contact: Sanchez D.O.
Instituto de Investigaciones Biotecnologicas (Univ. Nac. de Gral
San Martin)
Av. Gral Paz entre Albarillos y Constituyentes, INTI edificio 24
CP(1650) San Martin, Prov. de BS AS. Argentina
Tel: 54-11-4580-7255 ext 309
Fax: 54-11-4752-9639
Email: dsanchez@iib.unsam.edu.ar
Seq primer: F7
Class: shotgun.
FEATURES
Location/Qualifiers
1..481
/organism="Brucella melitensis biovar Abortus"
/strain="S-2308"
/db_xref="taxon:235"
/clone="B4G18"
/clone_lib="Sheared genomic library"
/note="Vector: pBluescript SK(-) (STRATAGENE); Genomic DNA
was mechanically sheared, blunt ended, and
size fractionated by agarose gel electrophoresis.
Fragments between 1.5-3 Kb were recovered and ligated to
the EcoRV site of the pBluescript SK (-) vector."
BASE COUNT
107 a 136 c 130 g 105 t 3 others
ORIGIN
Query Match 2.1%; Score 86.6; DB 17; Length 481;
Best Local Similarity 54.1%; Pred. No. 1.3e-11;
Matches 216; Conservative 0; Mismatches 181; Indels 2; Gaps 2;
QY 1079 GGCATTACGGTCATGCTCTGTGTAACCTGCTCATACAGGCAATGCGCCGCAAGGC 1138
Db ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
50 GCGCTTACCGTCTTCTCTGTGTAATCTGCTGCCAATGCGGAACCTGCAAAACGT 109
QY 1139 CTTCTCGGTATGCGAAGCGCGCATCCGGAGTTGGCGAAGTGATGAGGAACGC 1198
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
110 GATCTGCGATTTGCGGGATGACGACACAGATCTGCGCATTCATTGAANAACGCT 169
QY 1199 GACCTCCCGACGAATGGTTGATCGATCAACCCGACCGTTTCGGCGGAATCGCAA 1258
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
170 TGCCCTTCGCGTCAACAATGGTGGATCGCATTTGTACCGCAACGACAGATGAGACCGCAT 229
QY 1259 GAAGCTCAACGGCGCCAGTG-GGCTGGATGAGCACTGCCGCTGGTGGCGGAGATTCC 1317
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
230 ATCCATCAGCAGGACGTGAAGTTCTGGATCAATGCCCAATCACCGAGCCTTTCT 289
QY 1318 ATCACTGGGTGCTGGAAGACCAAGTTTGGCGATGGCGTCC-GCCGCTTCAAAAACCGGC 1376
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
290 CGCAATGGTCAATTGAAGACCGCTTTCCGACCGCGCGCCCTGCCCTGGGGAATTCCTGCG 349
QY 1377 GTGCAGATGGTCGGGAGCTGACGACCTGGGAGTACGTACAGATCCGATGCTCAATGCA 1436
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
350 GCAACCTTCGTGAGGATGTAGCAGCTTTTGAATTTGATGAACCTGCGCTGCTCAATGGC 409
QY 1437 GGGCATGTCATGCTCTGCTTCCAGCGATCTGTGTCGC 1475
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
410 AGCCATTCCACNCTTGCCCTATCTGGGTTTCTTGGCAGGC 448

RESULT 7
AW671891
LOCUS
DEFINITION Lgl_352_G02.b1_A002 Light Grown 1 (LGI) Sorghum bicolor mRNA EST 19-JUL-2000
624 bp linear

```

```

sequence.
AW671891
AW671891.1 GI:7535792
EST.
SOURCE
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 624)
Cordonnier-Pratt M.-M., Gingle,A., Marsala,C. and Pratt,L.H.
An EST database from Sorghum: light-grown seedlings
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: JEN REV
High quality sequence stop: 609
POLYA=No.
FEATURES
Location/Qualifiers
1..624
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Light Grown 1 (LGI)"
/note="Organ: 10- to 14-day-old light-grown (greenhouse)
seedlings; Vector: Lambda Zap; Site.1: XhoI; Site.2: EcoRI
; The library was made from poly-A RNA in the cloning
vector lambda ZAP II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT
146 a 217 c 135 g 126 t
ORIGIN
Query Match 2.1%; Score 85.4; DB 10; Length 624;
Best Local Similarity 55.6%; Pred. No. 3e-11;
Matches 233; Conservative 0; Mismatches 171; Indels 15; Gaps 3;
QY 585 CCCCCTATGATCGACGGGATCAAGCCTGGGATCGTGATTCGTTGTTAGTAACATT 644
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 CCTACTACGACCGTTCTAGCATCAAGCATGATCGTCCACGTTGTTGTTGGTTC 273
QY 645 TTTCAGCCCATGAGGCGTTCTACGTCGACGAGATTCTTGACACGCTCC-----GGAC 698
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
274 CACGTCGCCACTTGGCTGTCTACGTCGACAGCCTTCTTGAGCAGCTCAAGCTCAAGGAC 333
QY 699 TGGGCGATTGTTGGTGGTCTGACGGCGAGTTCACAAAGAAAAAGCCGAGAA 758
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 TGGGCTATCTCGGTTGGCATGAGTCTTCGATGCTGAGATGAGGATGCC----- 388
QY 759 TTCAGGCCCCAGGACTGCTGTTATTCCTGACCGAGAGCGCTCCGCGGAAGACGACG 818
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
389 -TCGTTCCCGAGGACAACCTCTACACATCATCGAGCGTTCCGCTCCCGGATC---CAAG 444
QY 819 GTGCGCTCATGTTGGGCGCTCGGTGACTATCTGTTGCCCGCGCGGATCCCGAAGCCGTG 878
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
445 GCTCGTCTGTCGCGGAGTATCACTGACTACATCTTCGCCCTTGACAACTGAGGCTGTC 504
QY 879 CTGAAGCATGTTGTTGATCCGCGCATCCGTTTCCATGACGATCCGGAAGCGGC 938
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
505 ATCGCCAAGATGGCTCACCCCGACACTCACATCGTCTCCATGACTATCACAGAGCGGG 564
QY 939 TACACATCAACGACGACCGGTCGTTTCGATCTGGAGATGCGGCGAGTAAAGCCGA 997
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
565 TACTTCTACACGAGACACACACAGCTCATGCCGAGACCCCGACATGCGCCCGCA 623

RESULT 8

```

BI191648
LOCUS BI191648 519 bp mRNA linear EST 10-JUL-2001
DEFINITION l1e03fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA library Fusarium sporotrichioides cDNA clone l1e03fs 5', mRNA sequence.
ACCESSION BI191648
VERSION BI191648.1 GI:14665327
KEYWORDS EST.
SOURCE Fusarium sporotrichioides.
ORGANISM Fusarium sporotrichioides.
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Mitosporic Hypocreales; Fusarium.
AUTHORS Ren, Q., Tag, A., Peplow, A., Lai, H., Kupfer, C., Peterson, A., Beremand, M. and Roe, B.
TITLE Analysis of a Fusarium sporotrichioides EST database
JOURNAL Unpublished (2001)
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 Contact Dr. Marian Beremand regarding clone availability Included
 is the best homolog from a blastx search of Genbank nr 04-09-01
 455 5e-45 gi|805056|emb|CAA604 (X86790) N3810 [Saccharomyces cerevisiae]
 Seq primer: T3
 High quality sequence stop: 423.
FEATURES Location/Qualifiers
 1..519
 /organism="Fusarium sporotrichioides"
 /strain="Tri 10"
 /db_xref="taxon:5514"
 /clone="l1e03fs"
 /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed cDNA library"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript XhoI; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT 137 a 166 c 116 g 100 t
ORIGIN
 Query Match 2.0%; Score 80.4; DB 13; Length 519;
 Best Local Similarity 50.9%; Pred. No. 5.7e-10;
 Matches 217; Conservative 0; Mismatches 206; Indels 3; Gaps 1;
 QY 576 GTCCAGGTCCTCCCTATGACATCGACGGGATCAAGCCTGGGATCGTGCATTCGGTGTA 635
 DB 97 GTCCAGGTCCTCCCTATGACATCGACGGGATCAAGCCTGGGATCGTGCATTCGGTGTA 156
 QY 636 GGTAACTTTTTCGAGCCCATGAGCGGTTCTACGTCGAGCAGATTCTTGAACACCGCTCGG 695
 DB 217 GAGCGGATGGGCCATCTGCGGTTATCGGCTCTCTAAGATGCTGCCATGCCGAT 276
 QY 756 GAATTCAGGCGCCAGACTGCTGTATTCCTGACCGAGAGCGCTCCGTCGGCAAGAGC 815
 DB 277 GTCTCTCAAGCCCAAGACCTCTACACCGTAAATTGAGCGATCCGCAAGGCGAGTTTC 336
 QY 816 ACGGTGCGGTGATGCGGCTGCGTGCATCTATCTGTTCGCCGCGGCTCCGGAAGCC 875
 DB 337 GCCGA---CGTTGAGGAGATATACCTCTCTTCAATTTTCGCTCCCGACGACCGTGAGGCC 393
 QY 876 GTGCTGAAGCACTTTGTTGATCGGCCCATCCGCGATCGTGTTCATGACGATCACGGAAGCC 935
 DB 394 GTCAATTGAAGATGGCTCATCCCGACACTACATCGTCTCTCTCATTACTAGAGC 453
 QY 936 GGCTACACATCAACGAGACGACCGGTGCGTTGATCTGGAGATGCGGCAGTAAAGGCC 995

DB 454 GGCTACTACTACAGAGAACACCCATCAATTGAAGGATGAGCCCGGACATCCAGCAC 513
 QY 996 GACCTC 1001
 DB 514 GATCTC 519
RESULT 9
LOCUS BI191768 519 bp mRNA linear EST 10-JUL-2001
DEFINITION l2e03fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA library Fusarium sporotrichioides cDNA clone l2e03fs 5', mRNA sequence.
ACCESSION BI191768
VERSION BI191768.1 GI:14665447
KEYWORDS EST.
SOURCE Fusarium sporotrichioides.
ORGANISM Fusarium sporotrichioides.
REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreales; Mitosporic Hypocreales; Fusarium.
AUTHORS Ren, Q., Tag, A., Peplow, A., Lai, H., Kupfer, C., Peterson, A., Beremand, M. and Roe, B.
TITLE Analysis of a Fusarium sporotrichioides EST database
JOURNAL Unpublished (2001)
COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 Contact Dr. Marian Beremand regarding clone availability Included
 is the best homolog from a blastx search of Genbank nr 04-09-01
 455 5e-45 gi|805056|emb|CAA604 (X86790) N3810 [Saccharomyces cerevisiae]
 Seq primer: T3
 High quality sequence stop: 382.
FEATURES Location/Qualifiers
 1..519
 /organism="Fusarium sporotrichioides"
 /strain="Tri 10"
 /db_xref="taxon:5514"
 /clone="l2e03fs"
 /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed cDNA library"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript XhoI; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT 137 a 166 c 116 g 100 t
ORIGIN
 Query Match 2.0%; Score 80.4; DB 13; Length 519;
 Best Local Similarity 50.9%; Pred. No. 5.7e-10;
 Matches 217; Conservative 0; Mismatches 206; Indels 3; Gaps 1;
 QY 576 GTCCAGGTCCTCCCTATGACATCGACGGGATCAAGCCTGGGATCGTGCATTCGGTGTA 635
 DB 97 GTCCAGGTCCTCCCTATGACATCGACGGGATCAAGCCTGGGATCGTGCATTCGGTGTA 156
 QY 636 GGTAACTTTTTCGAGCCCATGAGCGGTTCTACGTCGAGCAGATTCTTGAACACCGCTCGG 695
 DB 157 GCGGTTTCCACAGAGCTCATCTGCTATGTATGTCGACAAAGCTCTCGAAGATCGT 216
 QY 696 GACTGGCGATTTGTTGGTGTGGCTGACGGCAGTGCCTTCAAGAAAAACCGAG 755
 DB 217 GAGCGGATGGGCCATCTGCGGTTATCGGCTCTCTAAGATGCTGCCATGCCGAT 276
 QY 756 GAATTCAGGCGCCAGACTGCTGTATTCCTGACCGAGAGCGCTCCGTCGGCAAGAGC 815
 DB 277 GTCTCTCAAGCCCAAGACCTCTACACCGTAAATTGAGCGATCCGCAAGGCGAGTTTC 336
 QY 816 ACGGTGCGGTGATGCGGCTGCGTGCATCTATCTGTTCGCCGCGGCTCCGGAAGCC 875
 DB 337 GCCGA---CGTTGAGGAGATATACCTCTCTTCAATTTTCGCTCCCGACGACCGTGAGGCC 393
 QY 876 GTGCTGAAGCACTTTGTTGATCGGCCCATCCGCGATCGTGTTCATGACGATCACGGAAGCC 935
 DB 394 GTCAATTGAAGATGGCTCATCCCGACACTACATCGTCTCTCTCATTACTAGAGC 453
 DB 936 GGCTACACATCAACGAGACGACCGGTGCGTTGATCTGGAGATGCGGCAGTAAAGGCC 336

genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mfn-3xha/lacZ minitransposon containing lacZ, URA3, and tet resistance."

```
BASE COUNT      146 a   129 c   126 g   155 t
ORIGIN

Query Match      1.7%; Score 69.4; DB 17; Length 556;
Best Local Similarity 52.2%; Pred. No. 4.8e-07;
Matches 179; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

Qy 752 CGAGGAATTAAGCCGAGGACTGCTGTTATTCCTGACCGAGACGGCTCGTCCGGCAA 811
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 404 CGATGCCATGAAGCCCAAGATTGCTTATACACCTTGTGGAGCGGTGATCAAGGACAC 345
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 812 GAGCAGCGTGGCGCTGATGCGCGCTGCGTGACTATCTGCTGCCCGCGGATCCGGA 871
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 344 TAAGCTTATATCTGTC---GGTTCATTACTGCTTATGATGACCTCCGATATCCAG 288
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 872 AGCGGTGCTGAAGCATCTTGTGATCGCGGCATCCGATCGTTTCCATGACGATCACGGA 931
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 287 AGCTGTTATTGAAAGATGGCAATCCAGACACACATTTGTTTGTGACGGTTCACAGA 228
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 932 AGGCGGTACAAATCAACAGACGACCGGTGCTGCTGATCTGGAGATCGGCGAGTAAA 991
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 AAACGGTTACTACACAGTGAAGCAACAACCTCCTTAATGACAGATGCTCCCGAGATTAT 168
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 992 GGCGGACCTCAAGAACCGGAAACCGCTCTACCGTTTTTCGGTTACGTGGTFCGAGGCCCT 1051
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 167 CAATGATTGAACACCCAGAACCCAGATCTCTGTATGGTACCTATATGAGGCCCT 108
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1052 CGCTGCTGTTGGATGCCGTGTAAGGCAATTTACGGTCAATG 1094
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 107 GTTGTTGGTTACAAGAGAGCTCTTACCCCATTCACATTATTAGG 65
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
BG608411
LOCUS      BG608411          506 bp      mRNA      linear      EST 17-APR-2001
DEFINITION 296895 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  BG608411
VERSION    BG608411.1  GI:13658389
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 506)
AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL   Unpublished (2000)
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithemail.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGCAGG
Plate: 1 row: A column: 22
Seq primer: ATTTAGGTGACACTATG.
Location/Qualifiers
            1..506
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /clone_lib="MARC 1P1G"

genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mfn-3xha/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT      146 a   129 c   126 g   155 t
ORIGIN

Query Match      1.7%; Score 69.4; DB 17; Length 556;
Best Local Similarity 52.2%; Pred. No. 4.8e-07;
Matches 179; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

Qy 752 CGAGGAATTAAGCCGAGGACTGCTGTTATTCCTGACCGAGACGGCTCGTCCGGCAA 811
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 404 CGATGCCATGAAGCCCAAGATTGCTTATACACCTTGTGGAGCGGTGATCAAGGACAC 345
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 812 GAGCAGCGTGGCGCTGATGCGCGCTGCGTGACTATCTGCTGCCCGCGGATCCGGA 871
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 344 TAAGCTTATATCTGTC---GGTTCATTACTGCTTATGATGACCTCCGATATCCAG 288
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 872 AGCGGTGCTGAAGCATCTTGTGATCGCGGCATCCGATCGTTTCCATGACGATCACGGA 931
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 287 AGCTGTTATTGAAAGATGGCAATCCAGACACACATTTGTTTGTGACGGTTCACAGA 228
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 932 AGGCGGTACAAATCAACAGACGACCGGTGCTGCTGATCTGGAGATCGGCGAGTAAA 991
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 227 AAACGGTTACTACACAGTGAAGCAACAACCTCCTTAATGACAGATGCTCCCGAGATTAT 168
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 992 GGCGGACCTCAAGAACCGGAAACCGCTCTACCGTTTTTCGGTTACGTGGTFCGAGGCCCT 1051
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 167 CAATGATTGAACACCCAGAACCCAGATCTCTGTATGGTACCTATATGAGGCCCT 108
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1052 CGCTGCTGTTGGATGCCGTGTAAGGCAATTTACGGTCAATG 1094
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 107 GTTGTTGGTTACAAGAGAGCTCTTACCCCATTCACATTATTAGG 65
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
BG608411
LOCUS      BG608411          506 bp      mRNA      linear      EST 17-APR-2001
DEFINITION 296895 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  BG608411
VERSION    BG608411.1  GI:13658389
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 506)
AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL   Unpublished (2000)
COMMENT   Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithemail.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGCAGG
Plate: 1 row: A column: 22
Seq primer: ATTTAGGTGACACTATG.
Location/Qualifiers
            1..506
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /clone_lib="MARC 1P1G"

/tissue_type="pooled"
/lab_host="DH10B"
/note="vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT      130 a   118 c   138 g   120 t
ORIGIN

Query Match      1.3%; Score 54.2; DB 12; Length 506;
Best Local Similarity 49.8%; Pred. No. 0.0048;
Matches 137; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 1241 TTGCGCGGAATGCCCAAGAGCTCAACGGCGGCAGTGGGTGATGATGACGACCTGCGCT 1300
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 TGCACCCACAACTACTGTTGGAATACGACGACATCTGGGGTGAATGATCCCTCGCGCAT 61
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1301 GGTGCGCGAGGATTTCCATCAGTGGGTGGAAGACCACTTTTCGCGATGCGCTCCGCC 1360
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 TAGTTCGGAACCGTTTATCCAGTGGGTGGTGAAGATAACTTTCATCGCTGGCGCTCTGC 121
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1361 GCTTGAAGAAAGCCGCGCTGCAGATGTCGGGAGCGTGACGACGTGGAGTACGTCAAGAT 1420
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 122 CTGGGAAGTCGCAGATCTACAAATGGTGAATGATGTCCTGCCATGGGAAGATGAAACT 181
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1421 CCGAATGCTCAATGACGGGATGTCATGCTGCTGCTCCAGGCATTTCTGGTCCGCTATGA 1480
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 GCGGATGCTTAATGGCAGCCACTCTTTTCTCGCTTATCTGGGTTACCTCTCAGGATTGCG 241
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 1481 GAATGTGGATGACGCCATTTGAAGACAGCACTCC 1515
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 242 CCATATCAGTATTGCATGAGGATCGCGCATTTTC 276
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
AZ933800
LOCUS      BJ_Ba0001G08f B. japonicum BAC library Bradyrhizobium japonicum
DEFINITION genomic, DNA sequence.
ACCESSION  AZ933800
VERSION    AZ933800.1  GI:13775860
KEYWORDS   GSS.
SOURCE     Bradyrhizobium japonicum.
ORGANISM   Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE  1 (bases 1 to 788)
AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Golcoeschea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
            A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
            Genome Res. 11 (8), 1434-1440 (2001)
JOURNAL   21376150
MEDLINE   Contact: Wing RA
COMMENT   Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 725.
            Location/Qualifiers
                1..788
                /organism="Bradyrhizobium japonicum"
                /strain="USD110"
                /db_xref="taxon:375"
                /clone_lib="B. japonicum BAC library"
                /lab_host="E. coli"
                /note="vector: pindigo536; Site_1: HindIII"

BASE COUNT      128 a   271 c   258 g   131 t
ORIGIN

Query Match      1.2%; Score 51.4; DB 17; Length 788;
```


Db	989	BTBBSBSBSSSSSTSSSBBTSSBTSBSBSSTSSASBSBSSTSSASBSBSSTSSSBSSTTTBTSBBB	930
Qy	2438	TTCCGCTGCGGGTGATCCCGAGAAATACATAGGCATCCGGCTCGTTTCCGCTGCGCGCGC	2497
Db	929	STSSSSGSSSSSBBSTBBSBTTTBTBTKSTSTSSSTSTBTTTBSBSSSSSSBS	870
Qy	2498	GATGTGTGCGCTTTCGGCCCGTCCCATGCTCCTGGCGCAGCCAAATAGCCCGTGAA	2557
Db	869	SSBTTTBSBSTSSYSSSBTCTSKSTBTSGTBTBMSKBSTSTSBTSGSSBCG	810
Qy	2558	CGCGCTGCAGAAATCATTTCTTCGGGTGAAGCTCGCGCTGGCGCGGCATCGGCAC	2617
Db	809	GSGCSTSGSCSBCBKSTSSGSTTTCGBCBSTGGCSCCCVCSCCTCTCGTSTCC	750
Qy	2618	GGCGCATACGAACAGTCCGCTCATGAGTTCACAGCGCGCGATATTATTCGCATAGG	2677
Db	749	SCSGTSGTGBKCYSTSSSASTSGSSCSSTSSSGTGMTGGCCGHTTKGTGTCCTCG	690
Qy	2678	CCTTGCCCATTTCCGGGCATATC	2703
Db	689	GBYCCYBCCMTMCCGGSSCCSCSYC	664

[illegible]

```

/lab_host="DH10B"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
technologies. Contact : Feng Liang Life technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      188 a      282 c      243 g      147 t      1 others
ORIGIN

Query Match      1.2%; Score 48; DB 9; Length 861;
Best Local Similarity 49.6%; Pred. No. 0.28;
Matches 123; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Qy      761  C A G G C C C A G G A C T G C C T G A T T C C C T G A C C G A G A C G G T C C G C G G C A A G A G A C A C G G T 820

Db      319  C A T T G T C A T G A C T C T G A G A C G G G G T C A C C C A C A G G G T C C C A T C A T C A G G G C T A C G C 378

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:34:43 : Search time 40 Seconds
(without alignments)
1615.664 Million cell updates/sec

Title: US-09-926-163B-2
Perfect score: 485
Sequence: 1 MITRETLKSLPANVQAPPD.....KVIVLRKIIRKGVKAIPA 485

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	ID	Description
1	485	100.0	485	21	AAB231172
2	52	10.7	485	20	AAV49913
3	10	2.1	10	21	AAB231173
4	9	1.9	503	22	AAG89872
5	9	1.9	503	22	AAV79410
6	8	1.6	8	20	AAV49914
7	8	1.6	8	20	AAV49915
8	8	1.6	103	23	ABP33539
9	8	1.6	163	21	AAV99387
10	8	1.6	163	22	AAU12423

11	8	1.6	163	22	AAB66136	Protein of the inv
12	8	1.6	163	22	AAB36671	Human secretory pr
13	8	1.6	163	23	AB955524	Human angiogenesis
14	8	1.6	163	23	AB84918	Human PRO1273 prot
15	8	1.6	163	23	AB84918	Human PRO1273 prot
16	8	1.6	164	20	AAU83683	Human PRO protein,
17	8	1.6	172	22	AAV45273	Human secreted pro
18	8	1.6	172	22	AB95653	Human testicular a
19	8	1.6	172	22	AAV45273	Human reproductiv
20	8	1.6	261	22	ABG60855	Drosophila melanog
21	8	1.6	580	23	AB95021	Lactococcus lactis
22	7	1.4	1006	23	AB93763	Herbicideally activ
23	7	1.4	43	22	AB43361	Peptide #10867 enc
24	7	1.4	43	22	AB43361	Human brain expres
25	7	1.4	69	21	AAG52675	Arabidopsis thalia
26	7	1.4	72	22	AAU59450	Propionibacterium
27	7	1.4	73	22	AB43223	Peptide #10729 enc
28	7	1.4	73	22	AB26304	Protein #8303 enco
29	7	1.4	73	22	AAV64131	Human brain expres
30	7	1.4	73	22	AAV76955	Human bone marrow
31	7	1.4	73	22	AAV37063	Peptide #7474 enco
32	7	1.4	73	23	ABG46062	Human peptide enco
33	7	1.4	98	22	AAU54359	Propionibacterium
34	7	1.4	98	22	AB43453	Peptide #10959 enc
35	7	1.4	98	22	AB26421	Protein #8420 enco
36	7	1.4	98	22	AAV64382	Human brain expres
37	7	1.4	98	22	AAV7204	Human bone marrow
38	7	1.4	98	22	AAV21137	Peptide #7571 enco
39	7	1.4	98	22	AAV37341	Peptide #11378 enc
40	7	1.4	98	23	ABG46215	Human peptide enco
41	7	1.4	106	23	ABP41635	Human ovarian anti
42	7	1.4	132	22	AAU22384	Human cardiovascular
43	7	1.4	136	21	AAV01518	Human secreted pro
44	7	1.4	142	22	AAV80178	Human protein SEQ
45	7	1.4	143	22	AAV90408	C glutamicum prote

ALIGNMENTS

RESULT 1
AAB231172
ID AAB231172 standard; Protein; 485 AA.
XX
AC AAB231172;
XX
DT 29-JAN-2001 (first entry)
XX
DE Gluconobacter oxydans D-sorbitol dehydrogenase (SLDH).
XX
DE D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;
KW L-ascorbic acid biosynthesis; vitamin C.
XX
OS Gluconobacter oxydans.
XX
OS WO200055329-A1.
PN
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-JP01608.
XX
PR 17-MAR-1999; 99JP-0072810.
PR 06-AUG-1999; 99JP-0224679.
XX
PA (FUJI) FUJISAWA PHARM CO LTD.
XX
PI Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
PI Yamashita M, Takata Y;
XX
DR WPI; 2000-587530/55.
DR N-PSDB; AAA97430.
XX

PT Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
PT culturing its gene-transformed host cells, useful for producing
PT L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
PT acid production
PS Claim 5; Page 58-60; 72pp; Japanese.
XX
CC The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
CC (SLDH; AAB23172) and to the gene encoding it (AAA97430). SLDH has a
CC molecular weight of about 54 kDa and catalyses the conversion of
CC D-sorbitol into L-sorbose using NADP+ as a co-enzyme. SLDH specifically
CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
CC act on xylitol, ribitol, inositol and glycerol. The invention also
CC encompasses expression vectors and host cells comprising the
CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
CC The invention further relates to a method for preparing L-sorbose by
CC contacting the recombinant SLDH with D-sorbitol; a process for producing
CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
CC dehydrogenase and/or sorbose dehydrogenase with L-sorbose; and a
CC process for preparing L-ascorbic acid or its alkaline earth metals salts
CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
CC L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
CC acid production. The present sequence represents the Gluconobacter
CC oxydans SLDH protein.
XX
SQ Sequence 485 AA;

Query Match 100.0%; Score 485; DB 21; Length 485;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MITRETKSLPANVQAPPYDIDGIRPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
DB 1 MITRETKSLPANVQAPPYDIDGIRPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
QY 61 GLUTSDRSKKKAEFEKAQDCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKHLVD 120
DB 61 GLUTSDRSKKKAEFEKAQDCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKHLVD 120
QY 121 PAIRIVSMITTEGGYNINETTGAFLDNAVKADLNKPEKSTVFYVVEALRRRWDAGG 180
DB 121 PAIRIVSMITTEGGYNINETTGAFLDNAVKADLNKPEKSTVFYVVEALRRRWDAGG 180
QY 181 KFTVWSCDNLRHNGNARKAFGLGAKARDELAKEWIEENATFPNGMVDRTPTVSAEIA 240
DB 181 KFTVWSCDNLRHNGNARKAFGLGAKARDELAKEWIEENATFPNGMVDRTPTVSAEIA 240
QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNA 300
DB 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNA 300
QY 301 GHVMLCFPGILGVENVDDAIEDSELLGNLKNLKNKDIPTLKAPSGMTLEGYRDSVISR 360
DB 301 GHVMLCFPGILGVENVDDAIEDSELLGNLKNLKNKDIPTLKAPSGMTLEGYRDSVISR 360
QY 361 FSNKAMSQDTRLIASDGGSKVQVFTWTRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
DB 361 FSNKAMSQDTRLIASDGGSKVQVFTWTRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
QY 421 GYSESSEPTYGDAENKLAADDFESSKLKLPADGHRDLDTSELDOKVIVLRKIIREKGVK 480
DB 421 GYSESSEPTYGDAENKLAADDFESSKLKLPADGHRDLDTSELDOKVIVLRKIIREKGVK 480
QY 481 AAIPA 485
DB 481 AAIPA 485

RESULT 2
AAAY49913
ID AAAY49913 standard; Protein: 485 AA.
XX
AC AAAY49913;

XX 27-JAN-2000 (first entry)
DT Gluconobacter suboxydans L-sorbose reductase protein sequence.
DE
DE Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
KW mutant.
KW Gluconobacter suboxydans.
OS
XX Gluconobacter suboxydans.
XX AU9920390-A.
PN
XX 23-SEP-1999.
PD
XX 11-MAR-1999; 99AU-0020390.
PF
XX 13-MAR-1998; 98EP-0104546.
PR
XX (HOFF) HOFFMANN LA ROCHE & CO AG F.
PA
PI Hoshino T, Tazoe M, Shinjoh M, Kon T;
XX WPI; 1999-579275/49.
DR N-PSDB; AAZ35672.
XX
XX Set of genetically modified mutants not containing L sorbose reductase
PT
XX Claim 8; Page 18-21; 33pp; English.
PS
XX The present invention describes a genetically engineered
CC microorganism derived from a microorganism belonging to the genus
CC Gluconobacter or Acetobacter which is characterised in that the
CC biological activity for reducing L-sorbose is substantially
CC nullified by gene recombination gene. The present sequence represents
CC Gluconobacter suboxydans L-sorbose reductase.
XX N.B. This patent is equivalent to the basic NO9901197 in week 199949.
XX
SQ Sequence 485 AA;

Query Match 10.7%; Score 52; DB 20; Length 485;
Best Local Similarity 100.0%; Pred. No. 3 6e-43;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 186 MSCDNLRHNGNARKAFGLGAKARDELAKEWIEENATFPNGMVDRTPTVSA 237
DB 186 MSCDNLRHNGNARKAFGLGAKARDELAKEWIEENATFPNGMVDRTPTVSA 237

RESULT 3
AAB23173
ID AAB23173 standard; peptide; 10 AA.
XX
XX AAB23173;
AC
XX
DT 29-JAN-2001 (first entry)
XX
XX Gluconobacter oxydans SLDH N-terminal sequence.
DE
XX D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;
KW L-ascorbic acid biosynthesis; vitamin C; N-terminus.
XX
XX Gluconobacter oxydans.
OS
XX WO200055329-A1.
PN
XX 21-SEP-2000.
PD
XX 16-MAR-2000; 2000WO-JP01608.
XX
XX 17-MAR-1999; 99JP-0072810.
PR
XX 06-AUG-1999; 99JP-0224679.
PR
XX

PA (FUJI) FUJISAWA PHARM CO LTD.
 XX Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
 PI Yamashita M, Takata Y;
 XX WPI: 2000-587530/55.
 XX Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
 PT culturing its gene-transformed host cells, useful for producing
 PT L-sorboside or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 PT acid production
 XX
 PS Example 2; Page 28; 72pp; Japanese.
 XX
 CC The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
 CC (SLDH; AAB331172) and to the gene encoding it (AAA97430). SLDH has a
 CC molecular weight of about 54 kDa and catalyses the conversion of
 CC D-sorbitol into L-sorboside using NADP+ as a co-enzyme. SLDH specifically
 CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
 CC act on xylitol, ribitol, inositol and glycerol. The invention also
 CC encompasses expression vectors and host cells comprising the
 CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
 CC The invention further relates to a method for preparing L-sorboside by
 CC contacting the recombinant SLDH with D-sorbitol; a process for producing
 CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
 CC dehydrogenase and/or sorbosone dehydrogenase with L-sorboside; and a
 CC process for preparing L-ascorbic acid or its alkaline earth metals salts
 CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
 CC L-sorboside or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 CC acid production. The present sequence represents the N-terminal sequence
 CC of the Gluconobacter oxydans SLDH protein obtained in an exemplification
 CC of the invention.
 XX
 SQ Sequence 10' AA;

Query Match 2.1%; Score 10; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.015;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MITRETLKSL 10
 DB 1 MITRETLKSL 10
 |||||

RESULT 4
 AAG89872
 ID AAG89872 standard; Protein: 503 AA.

AC AAG89872;

DT 26-SEP-2001 (first entry)

XX C glutamicum protein fragment SEQ ID NO: 3626.

DE Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoquchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.
 DR N-PSDB; AAB65091.
 XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX Claim 17; SEQ ID NO: 3626; 246pp + Sequence Listing; English.
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 503 AA;

Query Match 1.9%; Score 9; DB 22; Length 503;
 Best Local Similarity 100.0%; Pred. No. 4.5;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PGIVHFGVG 34
 DB 33 PGIVHFGVG 41
 |||||

RESULT 5

AAB79410

ID AAB79410 standard; Protein: 503 AA.

XX AAB79410;

XX 30-APR-2001 (first entry)

DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:336.

KW Corynebacterium glutamicum; carbon metabolism and energy production;

KW SMP protein; sugar metabolism and oxidative phosphorylation protein;

KW fine chemical production; organic acid; proteinogenic amino acid;

KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;

KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;

KW carboxylate; aromatic compound; vitamin; cofactor; polyketide; enzyme;

KW diagnosis; Corynebacterium diphtheriae; evolutionary study.

XX Corynebacterium glutamicum.

OS WO200100844-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-IB00943.

XX 25-JUN-1999; 99US-0141031.

PR 08-JUL-1999; 99DE-1031412.

PR 08-JUL-1999; 99DE-1031413.

PR 08-JUL-1999; 99DE-1031419.

PR 08-JUL-1999; 99DE-1031420.

PR 08-JUL-1999; 99DE-1031424.

PR 08-JUL-1999; 99DE-1031428.

PR 08-JUL-1999; 99DE-1031431.

PR 08-JUL-1999; 99DE-1031433.

PR 08-JUL-1999; 99DE-1031434.

PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99DE-1040765.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX (BADI) BASF AG.
 PA Pompejus M, Kröger B, Schroeder H, Zelder O, Haberhauer G;
 PI N-PSDB; AAF71527.
 DR WPI; 2001-061975/07.
 DR N-PSDB; AAF71527.
 XX
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 XX Claim 20; Page 615-616; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 XX Sequence 503 AA;
 Query Match 1.9%; Score 9; DB 22; Length 503;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 PGIVHFGVG 34
 DB 33 PGIVHFGVG 41
 RESULT 6
 AAY49914
 ID AAY49914 standard; peptide: 8 AA.
 XX
 XX AAY49914;
 XX
 DT 27-JAN-2000 (first entry)
 XX
 DE Gluconobacter suboxydans L-sorbose reductase peptide SEQ ID NO:3.

XX Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
 KW mutant.
 XX
 OS Gluconobacter suboxydans.
 PN AU9920390-A.
 PD 23-SEP-1999.
 XX
 PF 11-MAR-1999; 99AU-0020390.
 XX
 PR 13-MAR-1998; 98EP-0104546.
 XX
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 PI Hoshino T, Tazoe M, Shinjoh M, Kon T;
 DR WPI; 1999-579276/49.
 XX
 PT Set of genetically modified mutants not containing L sorbose reductase
 XX
 PS Example 3; Fig 2; 33pp; English.
 XX
 CC The present invention describes a genetically engineered
 CC microorganism derived from a microorganism belonging to the genus
 CC Gluconobacter or Acetobacter which is characterised in that the
 CC biological activity for reducing L-sorbose is substantially
 CC nullified by gene recombination. The present sequence represents
 CC a Gluconobacter suboxydans L-sorbose reductase peptide, used in
 CC an example from the present invention for generating degenerate
 CC PCR primers.
 CC N.B. This patent is equivalent to the basic NO9901197 in week 199949.
 XX
 XX Sequence 8 AA;
 SQ
 Query Match 1.6%; Score 8; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 MTITEGGY 135
 DB 1 MTITEGGY 8
 RESULT 7
 AAY49915
 ID AAY49915 standard; peptide: 8 AA.
 XX
 AC AAY49915;
 XX
 DT 27-JAN-2000 (first entry)
 XX
 DE Gluconobacter suboxydans L-sorbose reductase peptide SEQ ID NO:4.
 XX
 KW Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
 KW mutant.
 OS Gluconobacter suboxydans.
 PN AU9920390-A.
 PD 23-SEP-1999.
 XX
 PF 11-MAR-1999; 99AU-0020390.
 XX
 PR 13-MAR-1998; 98EP-0104546.
 XX
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 PI Hoshino T, Tazoe M, Shinjoh M, Kon T;
 DR WPI; 1999-579276/49.

PR 02-SEP-1998; 98US-0098843;
 PR 09-SEP-1998; 98US-0099536;
 PR 09-SEP-1998; 98US-0099596;
 PR 09-SEP-1998; 98US-0099596;
 PR 09-SEP-1998; 98US-0099602;
 PR 09-SEP-1998; 98US-0099647;
 PR 10-SEP-1998; 98US-0099741;
 PR 10-SEP-1998; 98US-0099754;
 PR 10-SEP-1998; 98US-0099763;
 PR 10-SEP-1998; 98US-0099792;
 PR 10-SEP-1998; 98US-0099808;
 PR 10-SEP-1998; 98US-0099812;
 PR 10-SEP-1998; 98US-0099815;
 PR 10-SEP-1998; 98US-0099816;
 PR 15-SEP-1998; 98US-0100385;
 PR 15-SEP-1998; 98US-0100388;
 PR 15-SEP-1998; 98US-0100390;
 PR 16-SEP-1998; 98US-0100584;
 PR 16-SEP-1998; 98US-0100627;
 PR 16-SEP-1998; 98US-0100661;
 PR 16-SEP-1998; 98US-0100662;
 PR 16-SEP-1998; 98US-0100664;
 PR 17-SEP-1998; 98US-0100683;
 PR 17-SEP-1998; 98US-0100684;
 PR 17-SEP-1998; 98US-0100710;
 PR 17-SEP-1998; 98US-0100711;
 PR 17-SEP-1998; 98US-0100919;
 PR 17-SEP-1998; 98US-0100930;
 PR 18-SEP-1998; 98US-0100848;
 PR 18-SEP-1998; 98US-0100849;
 PR 18-SEP-1998; 98US-0101014;
 PR 18-SEP-1998; 98US-0101068;
 PR 18-SEP-1998; 98US-0101071;
 PR 22-SEP-1998; 98US-0101279;
 PR 23-SEP-1998; 98US-0101471;
 PR 23-SEP-1998; 98US-0101472;
 PR 23-SEP-1998; 98US-0101474;
 PR 23-SEP-1998; 98US-0101475;
 PR 23-SEP-1998; 98US-0101476;
 PR 23-SEP-1998; 98US-0101477;
 PR 23-SEP-1998; 98US-0101477;
 PR 24-SEP-1998; 98US-0101738;
 PR 24-SEP-1998; 98US-0101741;
 PR 24-SEP-1998; 98US-0101743;
 PR 24-SEP-1998; 98US-0101915;
 PR 24-SEP-1998; 98US-0101916;
 PR 29-SEP-1998; 98US-0102207;
 PR 29-SEP-1998; 98US-0102240;
 PR 29-SEP-1998; 98US-0102307;
 PR 29-SEP-1998; 98US-0102330;
 PR 29-SEP-1998; 98US-0102331;
 PR 30-SEP-1998; 98US-0102484;
 PR 30-SEP-1998; 98US-0102487;
 PR 30-SEP-1998; 98US-0102570;
 PR 30-SEP-1998; 98US-0102571;
 PR 01-OCT-1998; 98US-0102684;
 PR 01-OCT-1998; 98US-0102687;
 PR 02-OCT-1998; 98US-0102965;
 PR 06-OCT-1998; 98US-0102965;
 PR 06-OCT-1998; 98US-0103258;
 PR 07-OCT-1998; 98US-0103449;
 PR 07-OCT-1998; 98US-0103314;
 PR 07-OCT-1998; 98US-0103315;
 PR 07-OCT-1998; 98US-0103328;
 PR 07-OCT-1998; 98US-0103395;
 PR 07-OCT-1998; 98US-0103396;
 PR 07-OCT-1998; 98US-0103401;
 PR 08-OCT-1998; 98US-0103633;
 PR 08-OCT-1998; 98US-0103678;
 PR 08-OCT-1998; 98US-0103679;
 PR 14-OCT-1998; 98US-0103711;
 PR 20-OCT-1998; 98US-0104257;
 PR 20-OCT-1998; 98US-0104987;
 PR 20-OCT-1998; 98US-0105000;

PR 20-OCT-1998; 98US-0105002;
 PR 21-OCT-1998; 98US-0105104;
 PR 22-OCT-1998; 98US-0105169;
 PR 22-OCT-1998; 98US-0105266;
 PR 26-OCT-1998; 98US-0105693;
 PR 26-OCT-1998; 98US-0105694;
 PR 27-OCT-1998; 98US-0105807;
 PR 27-OCT-1998; 98US-0105881;
 PR 27-OCT-1998; 98US-0105882;
 PR 27-OCT-1998; 98US-0106062;
 PR 28-OCT-1998; 98US-0106023;
 PR 28-OCT-1998; 98US-0106029;
 PR 28-OCT-1998; 98US-0106030;
 PR 28-OCT-1998; 98US-0106032;
 PR 28-OCT-1998; 98US-0106033;
 PR 28-OCT-1998; 98US-0106178;
 PR 29-OCT-1998; 98US-0106248;
 PR 29-OCT-1998; 98US-0106384;
 PR 29-OCT-1998; 98US-0108500;
 PR 30-OCT-1998; 98US-0106464;
 PR 03-NOV-1998; 98US-0106856;
 PR 03-NOV-1998; 98US-0106902;
 PR 03-NOV-1998; 98US-0106905;
 PR 03-NOV-1998; 98US-0106919;
 PR 03-NOV-1998; 98US-0106932;
 PR 03-NOV-1998; 98US-0106934;
 PR 10-NOV-1998; 98US-0107783;
 PR 17-NOV-1998; 98US-0108775;
 PR 17-NOV-1998; 98US-0108779;
 PR 17-NOV-1998; 98US-0108787;
 PR 17-NOV-1998; 98US-0108788;
 PR 17-NOV-1998; 98US-0108801;
 PR 17-NOV-1998; 98US-0108802;
 PR 17-NOV-1998; 98US-0108806;
 PR 17-NOV-1998; 98US-0108807;
 PR 17-NOV-1998; 98US-0108867;
 PR 17-NOV-1998; 98US-0108925;
 PR 18-NOV-1998; 98US-0108848;
 PR 18-NOV-1998; 98US-0108849;
 PR 18-NOV-1998; 98US-0108850;
 PR 18-NOV-1998; 98US-0108851;
 PR 18-NOV-1998; 98US-0108852;
 PR 18-NOV-1998; 98US-0108858;
 PR 18-NOV-1998; 98US-0108904;
 PR 18-NOV-1998; 98US-0108904;

(GETH) GENENTECH INC.

Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;

WPI; 2000-237871/20.

N-PSDB; AAA37069.

New mammalian DNA sequences encoding transmembrane, receptor or secreted PRO polypeptides, useful for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interactions

Claim 12; Fig 96; 773pp; English.

AAA37022 to AAA37144 encode the new isolated human transmembrane, receptor or secreted PRO polypeptides given in AA199340 to AA199462. The transmembrane and receptor PRO proteins can be used for screening of potential peptide or small molecule inhibitors of the relevant receptor/ligand interactions. The polypeptides and nucleotide sequences encoding then have various industrial applications, including uses as pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent PCR primers and hybridisation probes used in the isolation of the PRO polypeptides from the present invention.

Sequence 163 AA;

Query Match 1.6%; Score 8; DB 21; Length 163;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
 DB 136 LYSLTETA 143
 RESULT 10
 AAU12423
 ID AAU12423 standard; Protein; 163 AA.
 AC AAU12423;
 DT 24-OCT-2001 (first entry)
 XX Human PRO1273 polypeptide sequence.
 DE Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200140466-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000WO-US32678.
 XX
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 09-DEC-1999; 99US-0170262.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 30-DEC-1999; 99WO-US31243.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 11-FEB-2000; 2000WO-US00376.
 PR 18-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 22-FEB-2000; 2000WO-US04342.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Beresini M, DeForge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 WPI: 2001-408281/43.
 DR N-PSDB; AAS21495.
 XX
 XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 12; Fig 504; 813pp; English.
 XX

CC AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 163 AA;
 Query Match 1.6%; Score 8; DB 22; Length 163;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 81 LYSLTETA 88
 DB 136 LYSLTETA 143
 RESULT 11
 AAB66136
 ID AAB66136 standard; protein; 163 AA.
 XX
 AC AAB66136;
 XX
 DT 02-APR-2001 (first entry)
 XX Protein of the invention #48.
 DE
 XX Secreted; transmembrane; gene therapy.
 KW
 XX Unidentified.
 OS
 XX WO200078961-A1.
 PN
 XX 28-DEC-2000.
 PD
 XX 18-FEB-2000; 2000WO-US04342.
 PF
 XX 23-JUN-1999; 99US-0141037.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99WO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 16-DEC-1999; 99WO-US30095.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
 PI Watanabe CK, Williams PM, Wood WI;
 XX
 WPI: 2001-071395/08.
 DR
 XX Secreted and transmembrane proteins and nucleic acids designated PRO,
 PT

PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -

XX Claim 1; Fig 96; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.

XX Sequence 163 AA:

Query Match 1.6%; Score 8; DB 22; Length 163;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 81 LYSLETETA 88
Db 136 LYSLETETA 143
|||||||

RESULT 12
AAB36671
ID AAB36671 standard; Protein; 163 AA.

XX AAB36671;

XX 14-MAR-2001 (first entry)

XX Human secretory protein TGC-715 SEQ ID NO:11.

XX Human; secretory protein; cancer; immune disease; infectious disease;
KW lung function disorder; liver function disorder; antiinflammatory;
KW gastrointestinal disorder; cytostatic; haematopoietic; anticoagulant;
KW immunomodulatory; hepatotropic; cell proliferation-stimulant;
KW cell migratory agent; cell differentiation-inducer.

XX Homo sapiens.

XX WO200071581-A1.

XX 30-NOV-2000.

XX 19-MAY-2000; 2000WO-JP03221.

XX 20-MAY-1999; 99JP-0140229.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Itoh Y, Mogi S, Tanaka H, Ohkubo S, Ogi K;

XX WPI; 2001-032023/04.

XX N-PSDB; AAC90711.

XX Novel secretory protein and its salt with e.g. anti-cancer,
PT anti-inflammatory and hematopoietic, effects, applicable as drugs in
PT remedies and preventives to treat diseases like cancer and immune
PT diseases -

XX Claim 1; Page 92-93; 122pp; Japanese.

XX AAC90701 to AAC90715 encode the human secretory proteins given in
CC AAB36661 to AAB36675. The proteins can have cytostatic,
CC anti-inflammatory, haematopoietic, anti-coagulant, immunomodulatory and
CC hepatotropic activities, and can be used as cell migratory agents, cell
CC proliferation-stimulants and cell differentiation-inducers. The proteins
CC are useful in the treatment and prevention of diseases such as cancer,
CC lung function disorder, liver function disorder, gastrointestinal
CC disorder and immune diseases. AAC90716 to AAC90755 represent PCR primers
CC which are used in the exemplification of the present invention.

XX Sequence 163 AA:

Query Match 1.6%; Score 8; DB 22; Length 163;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 81 LYSLETETA 88
Db 136 LYSLETETA 143
|||||||

RESULT 13
ABB95524
ID ABB95524 standard; Protein; 163 AA.

XX ABB95524;

XX 19-JUL-2002 (first entry)

XX Human angiogenesis related protein PRO1273 SEQ ID NO: 204.

XX Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiant; cytostatic; antiangiogenic; hypotensive; vulnery;

XX antiarteriosclerotic.

XX Homo sapiens.

XX WO200208284-A2.

XX 31-JAN-2002.

XX 09-JUL-2001; 2001WO-US21735.

XX 20-JUL-2000; 2000US-219556P.

XX 25-JUL-2000; 2000US-220624P.

XX 25-JUL-2000; 2000US-220664P.

XX 28-JUL-2000; 2000WO-US20710.

XX 02-AUG-2000; 2000US-222695P.

XX 17-AUG-2000; 2000US-0643657.

XX 23-AUG-2000; 2000WO-US23522.

XX 24-AUG-2000; 2000WO-US23328.

XX 07-SEP-2000; 2000US-230978P.

XX 15-SEP-2000; 2000US-000000P.

XX 18-SEP-2000; 2000US-0664610.

XX 18-SEP-2000; 2000US-0665350.

XX 24-OCT-2000; 2000US-242922P.

XX 08-NOV-2000; 2000US-0709238.

XX 10-NOV-2000; 2000WO-US30952.

XX 01-DEC-2000; 2000WO-US32678.

XX 20-DEC-2000; 2000US-0747259.

XX 20-DEC-2000; 2000WO-US34956.

XX 22-JAN-2001; 2001US-0767609.

XX 28-FEB-2001; 2001US-0796498.

XX 28-FEB-2001; 2001WO-US06520.

XX 01-MAR-2001; 2001WO-US06666.

XX 09-MAR-2001; 2001US-0802706.

XX 14-MAR-2001; 2001US-0808689.

XX 22-MAR-2001; 2001US-0816744.

XX 05-APR-2001; 2001US-0828366.

XX 10-MAY-2001; 2001US-0854208.

XX 10-MAY-2001; 2001US-0854280.

XX 25-MAY-2001; 2001US-0866028.

XX 25-MAY-2001; 2001US-0866034.

XX 30-MAY-2001; 2001WO-US17092.

XX 30-MAY-2001; 2001US-0870574.

XX 01-JUN-2001; 2001WO-US17443.

XX 20-JUN-2001; 2001WO-US17800.

XX 28-JUN-2001; 2001WO-US19692.

XX 28-JUN-2001; 2001WO-US00000.

```

PA (GETH ) GENENTECH INC.
PA (BAKE/) BAKER K P.
PA (FERR/) FERRARA N.
PA (GERB/) GERBER H.
PA (GERR/) GERRITSEN M E.
PA (GODD/) GODDARD A.
PA (GODO/) GODOWSKI P J.
PA (GURN/) GURNEY A L.
PA (HILL/) HILLAN K J.
PA (MARS/) MARSTERS S A.
PA (PANJ/) PAN J.
PA (PAON/) PAONI N F.
PA (STEP/) STEPHAN J F.
PA (WATA/) WATANABE C K.
PA (WILL/) WILLIAMS P M.
PA (WOOD/) WOOD W I.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX
XX WPI; 2002-171999/22.
XX N-PSDB; ABL95662.
XX
XX One hundred and eighty seven nucleic acids encoding PRO polypeptides,
XX useful in diagnosis and treatment of cardiovascular (e.g. myocardial
XX infarction), endothelial or angiogenic disorders in a mammal -
XX
XX Claim 11; Fig 204; 567pp; English.
XX
XX The present invention provides the protein and coding sequences of human
XX PRO proteins. These are useful for treating or diagnosing a
XX cardiovascular, endothelial or angiogenic disorder, including cardiac
XX hypertrophy, trauma, cancer, age-related macular degeneration,
XX atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
XX angina, myocardial infarctions, thrombophlebitis, lymphangiitis, tumour
XX angiogenesis (such as breast carcinoma and liver carcinoma) and wound
XX healing. The present sequence is a PRO protein of the invention.
XX
XX Sequence 163 AA;

Query Match 1.6%; Score 8; DB 23; Length 163;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
Db 136 LYSLTETA 143
|||||||

RESULT 14
ABB84918
ID ABB84918 standard; Protein; 163 AA.
XX
XX AC ABB84918;
XX
XX 16-MAY-2002 (first entry)
XX
XX Human PRO1273 protein sequence SEQ ID NO:204.
XX
XX Human; angiogenesis; cardiac; cytostatic; antiangiogenic; hypotensive;
XX vulnerable; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
XX gene therapy; cardiovascular disorder; endothelial disorder; cancer;
XX angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
XX age-related macular degeneration; arterial restenosis; angina;
XX rheumatoid arthritis; myocardial infarction; thrombophlebitis;
XX lymphangiitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
XX wound healing; chromosome mapping; gene mapping.
XX
XX Homo sapiens.
XX
XX OS
XX PN
XX WO200200690-A2.
XX

```

Query Match 1.6%; Score 8; DB 23; Length 163;

```

Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
|||||||
Db 136 LYSLTETA 143

RESULT 15
AAU83683
ID AAU83683 standard; Protein; 163 AA.
XX
XX AAU83683;
AC
XX
XX
08-MAY-2002 (first entry)
DT
XX
XX
DE Human PRO protein, Seq ID No 184.
XX
KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
KW breast cancer; prostate tumour; rectal tumour; liver tumour;
KW pericyte cell proliferation; chondrocyte cell proliferation;
KW tumour necrosis factor-alpha.
XX
XX Homo sapiens.
OS
XX

```

cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or liver tumour. The PRO polypeptides are useful for stimulating the proliferation of, or gene expression, in pericyte cells, for stimulating the proliferation or differentiation of chondrocyte cells, for stimulating the release of tumour necrosis factor-alpha from human blood, for stimulating or inhibiting the proliferation of normal human dermal fibroblast cells. The PRO polypeptide may also be used as molecular weight markers and for tissue typing. The PRO nucleic acids have applications in molecular biology, including use as hybridisation probes, and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO protein sequences of the invention.

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:31:32 ; Search time 17 Seconds
(without alignments)
839.419 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 485

Sequence: 1 MITRETLKSLPANVQAPPDY.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.4	25	1 US-07-966-187-4	Sequence 4, Appli
2	7	1.4	25	1 US-08-371-121-3	Sequence 3, Appli
3	7	1.4	25	3 US-07-927-381-5	Sequence 5, Appli
4	7	1.4	126	1 US-08-202-389-10	Sequence 10, Appli
5	7	1.4	266	2 US-08-685-992-12	Sequence 12, Appli
6	7	1.4	266	2 US-09-144-925-12	Sequence 12, Appli
7	7	1.4	303	2 US-08-286-819A-23	Sequence 23, Appli
8	7	1.4	303	3 US-08-980-357-23	Sequence 23, Appli
9	7	1.4	366	1 US-08-481-377-6	Sequence 6, Appli
10	7	1.4	366	3 US-08-827-336-9	Sequence 9, Appli
11	7	1.4	366	3 US-09-153-733A-6	Sequence 6, Appli
12	7	1.4	366	4 US-09-389-705-6	Sequence 6, Appli
13	7	1.4	366	4 US-09-357-905-9	Sequence 9, Appli
14	7	1.4	366	5 PCT-US94-00666-6	Sequence 6, Appli
15	7	1.4	377	4 US-09-269-137-4	Sequence 4, Appli
16	7	1.4	401	2 US-08-741-327E-11	Sequence 11, Appli
17	7	1.4	401	4 US-09-064-935-11	Sequence 11, Appli
18	7	1.4	404	3 US-08-844-065-2	Sequence 2, Appli
19	7	1.4	424	2 US-08-920-634-2	Sequence 2, Appli
20	7	1.4	455	4 US-09-308-003-13	Sequence 13, Appli
21	7	1.4	492	3 US-08-724-466B-2	Sequence 2, Appli
22	7	1.4	492	4 US-08-882-164D-2	Sequence 2, Appli
23	7	1.4	497	3 US-08-724-466B-4	Sequence 4, Appli
24	7	1.4	497	4 US-08-882-164D-4	Sequence 4, Appli
25	7	1.4	497	4 US-08-882-164D-32	Sequence 32, Appli
26	7	1.4	593	1 US-08-202-389-12	Sequence 12, Appli
27	7	1.4	593	1 US-08-018-129-5	Sequence 5, Appli

28 7 1.4 593 2 US-08-448-250-5 Sequence 5, Appli
29 7 1.4 968 1 US-08-426-236-2 Sequence 2, Appli
30 7 1.4 1684 3 US-08-665-259-25 Sequence 25, Appli
31 7 1.4 1684 3 US-08-762-500-25 Sequence 25, Appli
32 7 1.4 1704 3 US-08-762-500-75 Sequence 75, Appli
33 7 1.4 2291 2 US-08-286-819A-29 Sequence 29, Appli
34 7 1.4 2291 3 US-08-980-357-29 Sequence 29, Appli
35 7 1.4 5087 4 US-09-144-085-1 Sequence 1, Appli
36 7 1.4 6095 4 US-09-144-085-2 Sequence 2, Appli
37 6 1.2 11 3 US-08-159-339A-1157 Sequence 1157, Ap
38 6 1.2 14 1 US-08-440-391-11 Sequence 11, Appli
39 6 1.2 14 2 US-08-908-507A-11 Sequence 11, Appli
40 6 1.2 14 4 US-09-236-385A-11 Sequence 11, Appli
41 6 1.2 14 5 PCT-US96-06122-11 Sequence 11, Appli
42 6 1.2 20 1 US-08-321-071A-19 Sequence 19, Appli
43 6 1.2 34 1 US-08-014-153D-34 Sequence 34, Appli
44 6 1.2 35 4 US-08-942-686-6 Sequence 6, Appli
45 6 1.2 43 1 US-07-998-003A-73 Sequence 73, Appli

ALIGNMENTS

RESULT 1

US-07-966-187-4
; Sequence 4, Application US/07966187
; Patent No. 5477001

GENERAL INFORMATION:

APPLICANT: SASS, Catherine
APPLICANT: LEGUAY, Jean-Jacques

APPLICANT: GRISON, Rene

APPLICANT: TOPPAN, Alain

TITLE OF INVENTION: Recombinant DNA coding for a novel

TITLE OF INVENTION: protein having beta 1,3 glucanase activity, bacteria

TITLE OF INVENTION: containing this DNA, transformed plant cells and plants.

NUMBER OF SEQUENCES: 26

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER

STREET: King Street Station, Suite 500, 1800 Diagonal

STREET: Road, PO Box 299

CITY: ALEXANDRIA

STATE: VIRGINIA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/966.187

FILING DATE: 19930125

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR92/00268

FILING DATE: 25-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 91/03588

FILING DATE: 25-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: SAXE, Bernhard D

REGISTRATION NUMBER: 28,665

REFERENCE/DOCKET NUMBER: 16781/381

TELEPHONE: (703) 836-9300

TELEFAX: (703) 683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-07-966-187-4

Query Match 1.4%; Score 7; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 88 AFSGKST 94
Db 2 AFSGKST 8

RESULT 2

US-08-371-121-3
; Sequence 3, Application US/08371121
; Patent No. 5652123
; GENERAL INFORMATION:
; APPLICANT: CAPUT, Daniel
; APPLICANT: FERRARA, Pascal
; APPLICANT: GUILLEMOT, Jean-Claude
; APPLICANT: LEPLATOIS, Pascal
; APPLICANT: MINTY, Adrian
; APPLICANT: KAGHAD, Mourad
; APPLICANT: LABIT-LE BOUTEILLER, Christine
; APPLICANT: MAGAZIN, Marilyn
; TITLE OF INVENTION: Protein having a cytokine type
; TITLE OF INVENTION: activity, recombinant DNA coding for this protein,
; TITLE OF INVENTION: transformed cells and microorganisms.
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/371.121
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/938.161
; FILING DATE: 30-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR92/00280
; FILING DATE: 27-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 00137
; FILING DATE: 08-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 03904
; FILING DATE: 29-MAR-1991

ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28.665
; REFERENCE/DOCKET NUMBER: 16781/383
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-371-121-3

Query Match 1.4%; Score 7; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 88 AFSGKST 94
Db 2 AFSGKST 8

RESULT 3

US-07-927-391-5
; Sequence 5, Application US/07927391
; Patent No. 6001649
; GENERAL INFORMATION:
; APPLICANT: CAPUT, Daniel
; APPLICANT: FERRARA, Pascal
; APPLICANT: MILOUX, Brigitte
; APPLICANT: MINTY, Adrian
; APPLICANT: VITA, Natalio
; TITLE OF INVENTION: Protein having a cytokin type
; TITLE OF INVENTION: activity, and recombinant DNA, expression vector and hosts
; TITLE OF INVENTION: for its preparation.
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: King Street Station, Suite 500, 1800 Diagonal
; STREET: Road, PO Box 299
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22313-0299

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/927.391
; FILING DATE: 19920929
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: SAXE, Bernhard D.
; REGISTRATION NUMBER: 28.665
; REFERENCE/DOCKET NUMBER: 16781/369
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149

INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-927-391-5

Query Match 1.4%; Score 7; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 88 AFSGKST 94
Db 2 AFSGKST 8

RESULT 4

US-08-202-389-10
; Sequence 10, Application US/08202389
; Patent No. 5536636
; GENERAL INFORMATION:
; APPLICANT: Freeman Jr., Robert M.
; APPLICANT: Plutsky, Jorge
; APPLICANT: Neel, Benjamin G.
; APPLICANT: Rosenberg, Robert D.
; TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE

;; TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
;; NUMBER OF SEQUENCES: 54
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;; STREET: Two Militia Drive
;; CITY: Lexington
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02173
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/202,389
;; FILING DATE: 28-FEB-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/983,926
;; FILING DATE: 01-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/829,141
;; FILING DATE: 31-JAN-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/721,112
;; FILING DATE: 26-JUN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: BIH92-05WA
;; TELEPHONE: (617) 861-6240
;; TELEFAX: (617) 861-9540
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 126 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-202-389-10

Query Match 1.4%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 IIREKGV 479
Db 113 IIREKGV 119

RESULT 5
US-08-685-992-12
;; Sequence 12, Application US/08685992
;; Patent No. 5912138
;; GENERAL INFORMATION:
;; APPLICANT: Tonks, Nicholas
;; APPLICANT: Flint, Andrew J.
;; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
;; TITLE OF INVENTION: TYROSINE PHOSPHATASES
;; NUMBER OF SEQUENCES: 36
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
;; STREET: Two Militia Drive
;; CITY: Lexington
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02173
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows 95
;; SOFTWARE: FastSEQ for Windows Version 2.0b

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/685,992
;; FILING DATE: 25-JUL-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: CSHL96-03
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 781-861-6240
;; TELEFAX: 781-861-9540
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 266 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-685-992-12

Query Match 1.4%; Score 7; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 IIREKGV 479
Db 215 IIREKGV 221

RESULT 6
US-09-144-925-12
;; Sequence 12, Application US/09144925
;; Patent No. 5951979
;; GENERAL INFORMATION:
;; APPLICANT: Tonks, Nicholas
;; APPLICANT: Flint, Andrew J.
;; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
;; TITLE OF INVENTION: TYROSINE PHOSPHATASES
;; NUMBER OF SEQUENCES: 36
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
;; STREET: Two Militia Drive
;; CITY: Lexington
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02421-4799
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows 95
;; SOFTWARE: FastSEQ for Windows Version 2.0b
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/144,925
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/685,992
;; FILING DATE: July 25, 1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: CSHL96-032
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 781-861-6240
;; TELEFAX: 781-861-9540
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 266 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-144-925-12

Query Match 1.4%; Score 7; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 IIREKGV 479
|||||||
DB 215 IIREKGV 221

RESULT 7
US-08-286-819A-23
Sequence 23, Application US/08286819A
Patent No. 5871910

GENERAL INFORMATION:
APPLICANT: ARTHUR, MICHEL
APPLICANT: DUKTA-MALEN, SYLVIE
APPLICANT: MOLINAS, CATHERINE
APPLICANT: COURVALIN, PATRICE
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYMERES, IN PARTICULAR
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286.819A
FILING DATE: 05-AUG-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/174.682
FILING DATE: 28-DEC-1993
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917.146
FILING DATE: 10-AUG-1992
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR/91/00855
FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9013579
FILING DATE: 31-OCT-1990
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5871910man F.
REGISTRATION NUMBER: 24.618
REFERENCE/DOCKET NUMBER: 660-060-0 PCT

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR

SEQUENCE CHARACTERISTICS:
INFORMATION FOR SEQ ID NO: 23:
LENGTH: 303 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-286-819A-23

Query Match 1.4%; Score 7; DB 2; Length 303;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 KWIEENA 221
|||||||
DB 185 KWIEENA 191

RESULT 8
US-08-980-357-23
Sequence 23, Application US/08980357
Patent No. 6013508

GENERAL INFORMATION:
APPLICANT: ARTHUR, MICHEL
APPLICANT: DUKTA-MALEN, SYLVIE
APPLICANT: MOLINAS, CATHERINE
APPLICANT: COURVALIN, PATRICE
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYMERES, IN PARTICULAR
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/980.357
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286.819
FILING DATE: 05-AUG-1994
APPLICATION NUMBER: US 08/174.682
FILING DATE: 28-DEC-1993

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917.146
FILING DATE: 10-AUG-1992
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR/91/00855
FILING DATE: 29-OCT-1991

PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9013579
FILING DATE: 31-OCT-1990
ATTORNEY/AGENT INFORMATION:

NAME: Oblon, No. 6013508man F.
REGISTRATION NUMBER: 24.618
REFERENCE/DOCKET NUMBER: 660-060-0 PCT

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR

SEQUENCE CHARACTERISTICS:
INFORMATION FOR SEQ ID NO: 23:
LENGTH: 303 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-980-357-23

Query Match 1.4%; Score 7; DB 3; Length 303;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 KWIEENA 221
DB 185 KWIEENA 191

RESULT 9

US-08-481-377-6
; Sequence 6, Application US/08481377
; Patent No. 5808007
; GENERAL INFORMATION:
; APPLICANT: JOHNS HOPKINS UNIVERSITY
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.377
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: PCT/US94/00666
; FILING DATE: 12-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WETHERELL, JR. Ph.D., JOHN R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: FD2279 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-481-377-6

Query Match 1.4%; Score 7; DB 1; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
DB 51 VLRKIIR 57

RESULT 10

US-08-827-336-9
; Sequence 9, Application US/08827336
; Patent No. 6004780
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL
; TITLE OF INVENTION: GROWTH FACTOR HTR36
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE

STATE: MARYLAND
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827.336
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDY, A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF230
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8512
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-827-336-9

Query Match 1.4%; Score 7; DB 3; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
DB 51 VLRKIIR 57

RESULT 11

US-09-153-733A-6
; Sequence 6, Application US/09153733A
; Patent No. 6025475
; GENERAL INFORMATION:
; APPLICANT: JOHNS HOPKINS UNIVERSITY
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/153.733A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/481.377
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: WETHERELL, JR. Ph.D., JOHN R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: FD2279 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-153-733A-6

Query Match 1.4%; Score 7; DB 3; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
Db 51 VLRKIIR 57
|||||||

RESULT 12
US-09-389-705-6
; Sequence 6, Application US/09389705
; Patent No. 6391565
; GENERAL INFORMATION:

; APPLICANT: JOHNS HOPKINS UNIVERSITY
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/389,705
; FILING DATE: 03-Sep-1999
; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/153,733
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: WETHERELL, JR. Ph.D., JOHN R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: FD2279 PCT

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-389-705-6
Query Match 1.4%; Score 7; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
Db 51 VLRKIIR 57
|||||||

RESULT 13
US-09-357-905-9
; Sequence 9, Application US/09357905
; Patent No. 6413933
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL

; TITLE OF INVENTION: GROWTH FACTOR HTER36
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MARYLAND
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/357,905
; FILING DATE: 21-Jul-1999
; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,336
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDY, A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF230

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8512
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:

US-09-357-905-9
Query Match 1.4%; Score 7; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
Db 51 VLRKIIR 57
|||||||

RESULT 14
PCT-US94-00666-6
; Sequence 6, Application PC/TUS9400666
; GENERAL INFORMATION:
; APPLICANT: JOHNS HOPKINS UNIVERSITY
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/00666
; FILING DATE: 12-JAN-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: WETHERELL, JR. Ph.D., JOHN R.
; REGISTRATION NUMBER: 31,678

REFERENCE/DOCKET NUMBER: FD2279 PCT
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-00666-6

Query Match 1.4%; Score 7; DB 5; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLKRIIR 475
Db 51 VLKRIIR 57

RESULT 15

US-09-269-137-4
; Sequence 4, Application US/09269137
; Patent No. 6413751
; GENERAL INFORMATION:
; APPLICANT: Shapiro, Lucy
; APPLICANT: Kahng, Lyn Sue
; APPLICANT: Benkovic, Stephen J
; APPLICANT: Wright, Rachel
; APPLICANT: Stephens, Craig
; APPLICANT: Berdis, Anthony
; APPLICANT: Lee, Irene
; TITLE OF INVENTION: DNA Adenine Methyltransferases and Uses Thereof
; FILE REFERENCE: 99-292-B
; CURRENT APPLICATION NUMBER: US/09/269,137
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Brucella abortus
US-09-269-137-4

Query Match 1.4%; Score 7; DB 4; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 470 LRKIIRE 476
Db 364 LRKIIRE 370

Search completed: May 8, 2003, 16:35:15
Job time : 40 secs

THIS PAGE BLANK (U6PTD)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:33:44 ; Search time 32 Seconds
(without alignments)
1394.765 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 485

Sequence: 1 MITRETLKSLPANVOAPYD.....KVIVLRKIIRKGVKAAIPA 485

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	9	1.9	503	US-09-738-626-3626
2	8	1.6	163	US-10-028-072-504
3	8	1.6	163	US-10-121-049-504
4	8	1.6	163	US-10-123-904-504
5	8	1.6	163	US-10-140-470-504
6	8	1.6	163	US-10-175-746-504
7	8	1.6	163	US-10-176-918-504
8	8	1.6	163	US-10-176-921-504
9	8	1.6	163	US-10-227-884-184
10	8	1.6	163	US-10-137-865-504
11	8	1.6	163	US-10-140-474-504
12	8	1.6	163	US-10-142-431-504
13	8	1.6	163	US-10-143-114-504
14	8	1.6	163	US-10-230-163-184
15	8	1.6	163	US-10-006-856A-158
16	8	1.6	163	US-10-140-002-504
17	8	1.6	163	US-10-142-419-504
18	8	1.6	163	US-10-218-631-184
19	8	1.6	163	US-10-230-338-184

20	8	1.6	163	9	US-10-123-262-504	Sequence 504, App
21	8	1.6	163	9	US-10-142-423-504	Sequence 504, App
22	8	1.6	163	9	US-10-230-414-184	Sequence 184, App
23	8	1.6	163	9	US-10-006-818A-158	Sequence 158, App
24	8	1.6	163	9	US-10-121-050-504	Sequence 504, App
25	8	1.6	163	9	US-10-141-755-504	Sequence 504, App
26	8	1.6	163	9	US-10-143-032-504	Sequence 504, App
27	8	1.6	163	9	US-10-015-393A-158	Sequence 158, App
28	8	1.6	163	9	US-10-123-108-504	Sequence 504, App
29	8	1.6	163	9	US-10-123-236-504	Sequence 504, App
30	8	1.6	163	9	US-10-123-281-504	Sequence 504, App
31	8	1.6	163	9	US-10-140-921-504	Sequence 504, App
32	8	1.6	163	9	US-10-140-928-504	Sequence 504, App
33	8	1.6	163	9	US-10-216-159A-184	Sequence 184, App
34	8	1.6	163	9	US-09-946-374-158	Sequence 158, App
35	8	1.6	163	9	US-10-012-121A-158	Sequence 158, App
36	8	1.6	163	9	US-10-015-869A-158	Sequence 158, App
37	8	1.6	163	9	US-10-121-045-504	Sequence 504, App
38	8	1.6	163	9	US-10-123-252-504	Sequence 504, App
39	8	1.6	163	9	US-10-123-903-504	Sequence 504, App
40	8	1.6	163	9	US-10-124-819-504	Sequence 504, App
41	8	1.6	163	9	US-10-124-822-504	Sequence 504, App
42	8	1.6	163	9	US-10-140-925-504	Sequence 504, App
43	8	1.6	163	9	US-10-160-498-504	Sequence 504, App
44	8	1.6	163	9	US-10-218-849-184	Sequence 184, App
45	8	1.6	163	9	US-10-227-873-184	Sequence 184, App

ALIGNMENTS

RESULT 1
US-09-738-626-3626
; Sequence 3626, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3626
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3626

Query Match 1.9%; Score 9; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 26 PCIVHFGVG 34
DB 33 PCIVHFGVG 41

RESULT 2
US-10-028-072-504
: Sequence 504, Application US/10028072
: Publication No. US20030004311A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Beresini, Maureen
: APPLICANT: DeForge, Laura
: APPLICANT: Desnoyers, Luc
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Sherwood, Steven
: APPLICANT: Smith, Victoria
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K
: APPLICANT: Wood, William
: APPLICANT: Zhang
: TITLE OF INVENTION:
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/10/028,072
: CURRENT FILING DATE: 2001-12-19
: PRIOR APPLICATION NUMBER: 60/049911
: PRIOR FILING DATE: 1997-06-18
: PRIOR APPLICATION NUMBER: 60/056974
: PRIOR FILING DATE: 1997-08-26
: PRIOR APPLICATION NUMBER: 60/059113
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059115
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059117
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059122
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059184
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059263
: PRIOR FILING DATE: 1997-09-18
: PRIOR APPLICATION NUMBER: 60/059352
: PRIOR FILING DATE: 1997-09-19
: PRIOR APPLICATION NUMBER: 60/059588
: PRIOR FILING DATE: 1997-09-19
: PRIOR APPLICATION NUMBER: 60/059836
: PRIOR FILING DATE: 1997-09-24
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/062285
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/062287
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/062814
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/062816
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063045
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063082
: PRIOR FILING DATE: 1997-10-31
: PRIOR APPLICATION NUMBER: 60/063127
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063327
: PRIOR FILING DATE: 1997-10-27
: PRIOR APPLICATION NUMBER: 60/063329
: PRIOR FILING DATE: 1997-10-27
: PRIOR APPLICATION NUMBER: 60/063550
: PRIOR FILING DATE: 1997-10-28
: PRIOR APPLICATION NUMBER: 60/063561
: PRIOR FILING DATE: 1997-10-28
: PRIOR APPLICATION NUMBER: 60/063704
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063733
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063735
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063738
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063755
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/064248
: PRIOR FILING DATE: 1997-11-03
: PRIOR APPLICATION NUMBER: 60/064809
: PRIOR FILING DATE: 1997-11-07
: PRIOR APPLICATION NUMBER: 60/065186
: PRIOR FILING DATE: 1997-11-12
: PRIOR APPLICATION NUMBER: 60/065846
: PRIOR FILING DATE: 1997-11-17
: PRIOR APPLICATION NUMBER: 60/066364
: PRIOR FILING DATE: 1997-11-21
: PRIOR APPLICATION NUMBER: 60/066453
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/066511
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/066770
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/069212
: PRIOR FILING DATE: 1997-12-11
: PRIOR APPLICATION NUMBER: 60/069278
: PRIOR FILING DATE: 1997-12-11
: PRIOR APPLICATION NUMBER: 60/069334
: PRIOR FILING DATE: 1997-12-11
: PRIOR APPLICATION NUMBER: 60/069694
: PRIOR FILING DATE: 1997-12-16
: PRIOR APPLICATION NUMBER: 60/072320
: PRIOR FILING DATE: 1998-01-23
: PRIOR APPLICATION NUMBER: 60/073612
: PRIOR FILING DATE: 1998-02-04
: PRIOR APPLICATION NUMBER: 60/074086
: PRIOR FILING DATE: 1998-02-09
: PRIOR APPLICATION NUMBER: 60/074092
: PRIOR FILING DATE: 1998-02-09
: PRIOR APPLICATION NUMBER: 60/077791
: PRIOR FILING DATE: 1998-03-12
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/079294
: PRIOR FILING DATE: 1998-03-25
: PRIOR APPLICATION NUMBER: 60/079663
: PRIOR FILING DATE: 1998-02-27
: PRIOR APPLICATION NUMBER: 60/079728
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/080165
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/081203
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081229
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081695
: PRIOR FILING DATE: 1998-04-14
: PRIOR APPLICATION NUMBER: 60/081817
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081818
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/082999
: PRIOR FILING DATE: 1998-04-24
: PRIOR APPLICATION NUMBER: 60/083322
: PRIOR FILING DATE: 1998-04-28
: PRIOR APPLICATION NUMBER: 60/083545
: PRIOR FILING DATE: 1998-04-29
: PRIOR APPLICATION NUMBER: 60/084600
: PRIOR FILING DATE: 1998-05-07
: PRIOR APPLICATION NUMBER: 60/084627
: PRIOR FILING DATE: 1998-05-07

;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085149
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/086414
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/086430
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087106
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088730
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088741
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090349
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090538
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
| | | | | | | |
Db 136 LYSLTETA 143

RESULT 3

US-10-121-049-504
; Sequence 504, Application US/10121049
; Publication No. US2003002239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen

;; APPLICANT: DeForge, Laura
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Sherwood, Steven
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3330R1C17
;; CURRENT APPLICATION NUMBER: US/10/121,049
;; CURRENT FILING DATE: 2002-04-12
;; Prior Application removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 504
;; LENGTH: 163
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-121-049-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
| | | | | | | |
Db 136 LYSLTETA 143

RESULT 4

US-10-123-904-504
; Sequence 504, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Beresini, Maureen
;; APPLICANT: DeForge, Laura
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Sherwood, Steven
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3330R1C54
;; CURRENT APPLICATION NUMBER: US/10/123,904
;; CURRENT FILING DATE: 2002-04-16
;; Prior Application removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 504
;; LENGTH: 163
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-123-904-504

Query Match

1.6%; Score 8; DB 9; Length 163;

Best Local Similarity 100.0%; Pred. No. 7.9; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLETETA 88
Db 136 LYSLETETA 143

RESULT 5

US-10-140-470-504
; Sequence 504, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C160
; CURRENT APPLICATION NUMBER: US/10/140.470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLETETA 88
Db 136 LYSLETETA 143

RESULT 6

US-10-175-746-504
; Sequence 504, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William

; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C353
; CURRENT APPLICATION NUMBER: US/10/175.746
; CURRENT FILING DATE: 2002-06-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLETETA 88
Db 136 LYSLETETA 143

RESULT 7

US-10-176-918-504
; Sequence 504, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C382
; CURRENT APPLICATION NUMBER: US/10/176.918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLETETA 88
Db 136 LYSLETETA 143

RESULT 8

US-10-176-921-504
; Sequence 504, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.


```

; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-921-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 9
US-10-227-884-184
; Sequence 184, Application US/10227884
; Publication No. US20030027988A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330PIC79
; CURRENT APPLICATION NUMBER: US/10/227,884
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 10/119,480
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089905
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090691
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/095318
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/095916
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/096146
; PRIOR FILING DATE: 1998-08-11
; PRIOR APPLICATION NUMBER: 60/096791
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 60/097986
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: 60/098544
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099803
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099811
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100038
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
```

```

; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: 60/101477
; PRIOR FILING DATE: 1998-09-23
; PRIOR APPLICATION NUMBER: 60/101738
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101741
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101786
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: 60/101916
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/101922
; PRIOR FILING DATE: 1998-09-24
; PRIOR APPLICATION NUMBER: 60/106178
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/106248
; PRIOR FILING DATE: 1998-10-29
; PRIOR APPLICATION NUMBER: 60/106464
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 60/106905
; PRIOR FILING DATE: 1998-11-03
; PRIOR APPLICATION NUMBER: 60/108787
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/108801
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/108849
; PRIOR FILING DATE: 1998-11-18
; PRIOR APPLICATION NUMBER: 60/112422
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/112966
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115558
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115565
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/115733
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/119549
; PRIOR FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 60/123618
; PRIOR FILING DATE: 1999-03-10
; PRIOR APPLICATION NUMBER: 60/125259
; PRIOR FILING DATE: 1999-03-19
; PRIOR APPLICATION NUMBER: 60/125775
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/126773
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: 60/127887
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/130232
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131022
; PRIOR FILING DATE: 1999-04-26
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131445
; PRIOR FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: 60/134287
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/140650
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/140723
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/141037
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/144758
; PRIOR FILING DATE: 1999-07-20

```

```

; PRIOR APPLICATION NUMBER: 60/145698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 60/146222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 60/146963
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/149320
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149638
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/151733
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/164418
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: 60/166361
; PRIOR FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/169445
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169835

```

Query Match 1.6%; Score 8; DB 9; Length 163;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSUTETA 88
 |||||
 Db 136 LYSUTETA 143

RESULT 10

```

US-10-137-865-504
; Sequence 504, Application US/10137865.
; Publication No. US20030032155A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Leura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C154
; CURRENT APPLICATION NUMBER: US/10/137,865
; CURRENT FILING DATE: 2002-05-03
; Prior Application removed - See Palm or File wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-137-865-504

```

Query Match 1.6%; Score 8; DB 9; Length 163;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSUTETA 88
 |||||
 Db 136 LYSUTETA 143

```
RESULT 11
US-10-140-474-504
; Sequence 504, Application US/10140474
; Publication No. US20030032156A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-474-504

Query Match          1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 12
US-10-142-431-504
; Sequence 504, Application US/10142431
; Publication No. US20030036179A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC251
; CURRENT APPLICATION NUMBER: US/10/142,431
; CURRENT FILING DATE: 2002-05-10
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-431-504

Query Match          1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 13
US-10-143-114-504
; Sequence 504, Application US/10143114
; Publication No. US20030036180A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC211
; CURRENT APPLICATION NUMBER: US/10/143,114
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-143-114-504

Query Match          1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 14
US-10-230-163-184
; Sequence 184, Application US/10230163
; Publication No. US20030036635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
```

```
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-431-504

Query Match          1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 13
US-10-143-114-504
; Sequence 504, Application US/10143114
; Publication No. US20030036180A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC211
; CURRENT APPLICATION NUMBER: US/10/143,114
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-143-114-504

Query Match          1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 14
US-10-230-163-184
; Sequence 184, Application US/10230163
; Publication No. US20030036635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
```

APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin L.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3530PTC96
CURRENT APPLICATION NUMBER: US/10/230,163
CURRENT FILING DATE: 2002-08-28
PRIOR APPLICATION NUMBER: 10/119,480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/0959113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/069873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089905
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090691
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/095302
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095318
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095916
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/096146
PRIOR FILING DATE: 1998-08-11
PRIOR APPLICATION NUMBER: 60/096791
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/097986
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098544
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100038
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101786
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101922
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/106178
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106248
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 60/106464
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: 60/106905
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/108787
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108801
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108849
PRIOR FILING DATE: 1998-11-18
PRIOR APPLICATION NUMBER: 60/112422
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113296
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119549
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/123618
PRIOR FILING DATE: 1999-03-10
PRIOR APPLICATION NUMBER: 60/125259
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 60/125775
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/126773

```
; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: 60/127887
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/130232
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131022
; PRIOR FILING DATE: 1999-04-26
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131445
; PRIOR FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: 60/134287
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/140650
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/140723
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/141037
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/144758
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/145698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 60/146222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 60/146963
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/149320
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149638
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/151733
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/164418
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: 60/166361
; PRIOR FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/169445
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169835

Query Match      1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 LYSLETETA 88
      |||||
Db      136 LYSLETETA 143

Search completed: May  8, 2003, 16:41:48
Job time : 34 secs

; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC59
; CURRENT APPLICATION NUMBER: US/10/140.002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-002-504

Query Match      1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      81 LYSLETETA 88
      |||||
Db      136 LYSLETETA 143

Search completed: May  8, 2003, 16:41:48
Job time : 34 secs

; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
```

THIS PAGE BLANK (USPTO)